Text as Data

Justin Grimmer

Associate Professor
Department of Political Science
Stanford University

November 18th, 2014

Final Project

Poster Session: 12/4

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable
- Less is More

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable
- Less is More

Papers Due: 12/12, 5pm This is a hard deadline!

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable
- Less is More

Papers Due: 12/12, 5pm **This is a hard deadline!**

- An academic paper $\rightsquigarrow$ appropriate for your field

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable
- Less is More

Papers Due: 12/12, 5pm **This is a hard deadline!**

- An academic paper $\rightsquigarrow$ appropriate for your field
  - Succinct (not cute) title

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable
- Less is More

Papers Due: 12/12, 5pm This is a hard deadline!

- An academic paper $\rightsquigarrow$ appropriate for your field
    - Succinct (not cute) title
    - Abstract that explains your contribution

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable
- Less is More

Papers Due: 12/12, 5pm **This is a hard deadline!**

- An academic paper$\rightsquigarrow$ appropriate for your field
  - Succinct (not cute) title
  - Abstract that explains your contribution
  - Introduction that explains why your paper exists

Final Project

Poster Session: 12/4

- Make posters in LaTeX or related software for asthetics
- Landscape orientation is preferable
- Less is More

Papers Due: 12/12, 5pm **This is a hard deadline!**

- An academic paper $\rightsquigarrow$ appropriate for your field
  - Succinct (not cute) title
  - Abstract that explains your contribution
  - Introduction that explains why your paper exists
  - Presentation of Results

Final Project

Poster Session: 12/4

- Make posters in $\LaTeX$ or related software for asthetics
- Landscape orientation is preferable
- Less is More

Papers Due: 12/12, 5pm **This is a hard deadline!**

- An academic paper $\rightsquigarrow$ appropriate for your field
  - Succinct (not cute) title
  - Abstract that explains your contribution
  - Introduction that explains why your paper exists
  - Presentation of Results
  - Avoid: Long/breezy lit reviews that merely list previous scholarship

Supervised Learning: Ensemble Learning

1) Task

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding$\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding$\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k|\boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$
   - Functional form f

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) = f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$
   - Functional form f
   - Necessarily requires consequential assumptions

Supervised Learning: Ensemble Learning

1) Task
    - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) = f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

- "Coefficients" $\boldsymbol{\beta}$
- Tuning parameters $\boldsymbol{\lambda}$
- Functional form f
- Necessarily requires consequential assumptions
- Ensembles: aggregate classifiers to increase performance

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$
   - Functional form f
   - Necessarily requires consequential assumptions
   - Ensembles: aggregate classifiers to increase performance
3) Optimization

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding$\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$
   - Functional form f
   - Necessarily requires consequential assumptions
   - Ensembles: aggregate classifiers to increase performance
3) Optimization
   - Optimization of individual classifiers$\rightsquigarrow$ methods already discussed

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$
   - Functional form f
   - Necessarily requires consequential assumptions
   - Ensembles: aggregate classifiers to increase performance
3) Optimization
   - Optimization of individual classifiers $\rightsquigarrow$ methods already discussed
   - Determination of weights on models $\rightsquigarrow$ equal (bagging), out of sample performance via cross validation (super learning)

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) = f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$
   - Functional form f
   - Necessarily requires consequential assumptions
   - Ensembles: aggregate classifiers to increase performance
3) Optimization
   - Optimization of individual classifiers $\rightsquigarrow$ methods already discussed
   - Determination of weights on models $\rightsquigarrow$ equal (bagging), out of sample performance via cross validation (super learning)
4) Validation

Supervised Learning: Ensemble Learning

1) Task
   - Subsidize hand coding $\rightsquigarrow$ learn a rule between labels and features
2) Objective function

$$\Pr(Y_i = C_k | \boldsymbol{x}_i) \quad = \quad f(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda})$$

   - "Coefficients" $\boldsymbol{\beta}$
   - Tuning parameters $\boldsymbol{\lambda}$
   - Functional form f
   - Necessarily requires consequential assumptions
   - Ensembles: aggregate classifiers to increase performance
3) Optimization
   - Optimization of individual classifiers $\rightsquigarrow$ methods already discussed
   - Determination of weights on models $\rightsquigarrow$ equal (bagging), out of sample performance via cross validation (super learning)
4) Validation
   - Out of sample predictive performance

Ensemble Learning: Intuition

Heuristic (upon which we'll improve):

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse$\rightarrow$ ensemble methods improve

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse $\rightarrow$ ensemble methods improve

Intuition:

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse$\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse$\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).
- True labels: evenly distributed across two categories

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse $\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).
- True labels: evenly distributed across two categories
- Three classifiers with 75% accuracy, but independent

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse $\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).
- True labels: evenly distributed across two categories
- Three classifiers with 75% accuracy, but independent
- Implement majority voting rule

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse$\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).
- True labels: evenly distributed across two categories
- Three classifiers with 75% accuracy, but independent
- Implement majority voting rule

$$\Pr(\text{Correct Guess}|\text{Votes})$$

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse$\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).
- True labels: evenly distributed across two categories
- Three classifiers with 75% accuracy, but independent
- Implement majority voting rule

$$\Pr(\text{Correct Guess}|\text{Votes}) = \Pr(3 \text{ correct}) + \Pr(2 \text{ correct})$$

Ensemble Learning: Intuition

Heuristic (upon which we'll improve): if classifiers are accurate and diverse$\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).
- True labels: evenly distributed across two categories
- Three classifiers with 75% accuracy, but independent
- Implement majority voting rule

$$\begin{aligned} \Pr(\text{Correct Guess}|\text{Votes}) &= \Pr(3 \text{ correct}) + \Pr(2 \text{ correct}) \\ &= 0.75^3 + 3 \times (0.75^2 \times 0.25) \end{aligned}$$

Ensemble Learning: Intuition

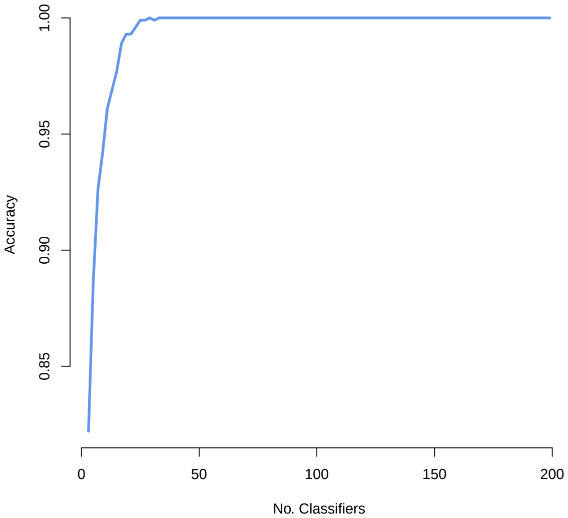
Heuristic (upon which we'll improve): if classifiers are accurate and diverse$\rightarrow$ ensemble methods improve

Intuition:

- Classify documents into two categories (Category 1, Category 2).
- True labels: evenly distributed across two categories
- Three classifiers with 75% accuracy, but independent
- Implement majority voting rule

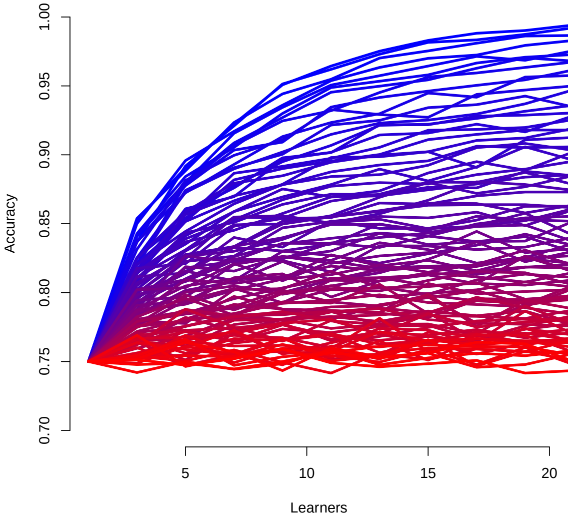
$$
\begin{aligned}
\Pr(\text{Correct Guess}|\text{Votes}) &= \Pr(3 \text{ correct}) + \Pr(2 \text{ correct}) \\
&= 0.75^3 + 3 \times (0.75^2 \times 0.25) \\
&= 0.844
\end{aligned}
$$

Ensemble Learning: Intuition



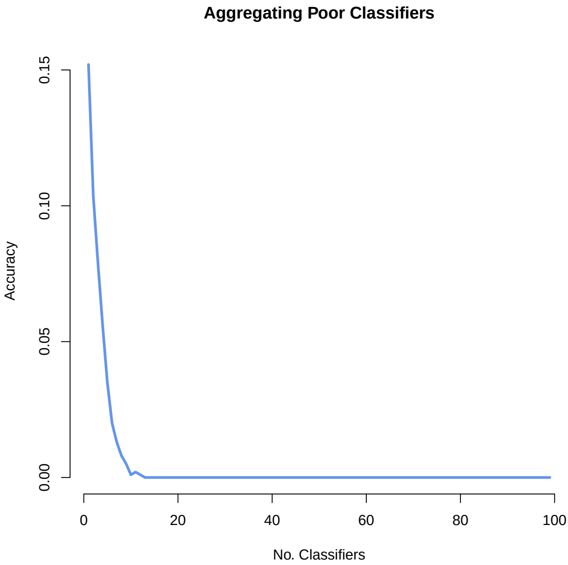
Ensemble Learning: Intuition

Diverse and Accurate matter.



Ensemble Learning: Intuition

Diverse and Accurate matter.



Wisdom of the Crowds:

Goal: estimate a document's category $\rightsquigarrow$ $Y \in \{0, 1\}$

Wisdom of the Crowds:

Goal: estimate a document's category $\rightsquigarrow$ $Y \in \{0,1\}$
Classifiers: (suppose) a sequence of identically distributed (not independent) random variables.

Wisdom of the Crowds:

Goal: estimate a document's category $\rightsquigarrow$ $Y \in \{0, 1\}$
Classifiers: (suppose) a sequence of identically distributed (not independent) random variables.
Suppose $Y = 1$

Wisdom of the Crowds:

Goal: estimate a document's category $\rightsquigarrow$ $Y \in \{0, 1\}$
Classifiers: (suppose) a sequence of identically distributed (not independent) random variables.
Suppose $Y = 1$
Guess from classifer m is B_m with $\Pr(B_i = 1) = p > 0.5$.

Wisdom of the Crowds:

Goal: estimate a document's category $\rightsquigarrow$ $Y \in \{0, 1\}$
Classifiers: (suppose) a sequence of identically distributed (not independent) random variables.
Suppose $Y = 1$
Guess from classifer m is B_m with $\Pr(B_i = 1) = p > 0.5$.

$$\bar{B} = \sum_{m=1}^{M} \frac{B_m}{M}$$

Wisdom of the Crowds:

Goal: estimate a document's category$\rightsquigarrow$ $Y \in \{0, 1\}$
Classifiers: (suppose) a sequence of identically distributed (not independent) random variables.
Suppose $Y = 1$
Guess from classifer m is B_m with $\Pr(B_i = 1) = p > 0.5$.

$$\bar{B} \quad = \quad \sum_{m=1}^{M} \frac{B_m}{M}$$

Wisdom of crows (Condorcet Jury Theorem)

Wisdom of the Crowds:

Goal: estimate a document's category$\rightsquigarrow Y \in \{0,1\}$
Classifiers: (suppose) a sequence of identically distributed (not independent) random variables.
Suppose $Y = 1$
Guess from classifer m is B_m with $\Pr(B_i = 1) = p > 0.5$.

$$\bar{B} = \sum_{m=1}^{M} \frac{B_m}{M}$$

Wisdom of crows (Condorcet Jury Theorem)

$$\lim_{M\to\infty} P(\bar{B} > 0.5) = 1$$

Wisdom of the Crowds

Suppose B_m have variance σ^2 and pairwise correlation ρ.

Wisdom of the Crowds

Suppose B_m have variance σ^2 and pairwise correlation ρ.
Then,

Wisdom of the Crowds

Suppose B_m have variance σ^2 and pairwise correlation ρ.
Then,

$$\text{var}(\bar{B})$$

Wisdom of the Crowds

Suppose B_m have variance σ^2 and pairwise correlation ρ.
Then,

$$\text{var}(\bar{B}) = \text{var}\left(\sum_{i=1}^{M} \frac{B_i}{M}\right)$$

Wisdom of the Crowds

Suppose B_m have variance σ^2 and pairwise correlation ρ.
Then,

$$\begin{aligned}
\text{var}(\bar{B}) &= \text{var}\left(\sum_{i=1}^{M} \frac{B_i}{M}\right) \\
&= \frac{1}{M^2}\sum_{i=1}^{M}\text{var}\left(B_i\right) + \frac{2}{M^2}\sum_{i<j}\text{cov}(B_i, B_j)
\end{aligned}$$

Wisdom of the Crowds

Suppose B_m have variance σ^2 and pairwise correlation ρ.
Then,

$$\begin{aligned}
\text{var}(\bar{B}) &= \text{var}\left(\sum_{i=1}^{M} \frac{B_i}{M}\right) \\
&= \frac{1}{M^2}\sum_{i=1}^{M} \text{var}\left(B_i\right) + \frac{2}{M^2}\sum_{i<j} \text{cov}(B_i, B_j) \\
&= \frac{M\sigma^2}{M^2} + \frac{2}{M^2}\rho\sigma^2\binom{M}{2}
\end{aligned}$$

Wisdom of the Crowds

Suppose B_m have variance σ^2 and pairwise correlation ρ.
Then,

$$
\begin{aligned}
\text{var}(\bar{B}) &= \text{var}\left(\sum_{i=1}^{M} \frac{B_i}{M}\right) \\
&= \frac{1}{M^2}\sum_{i=1}^{M}\text{var}\left(B_i\right) + \frac{2}{M^2}\sum_{i<j}\text{cov}(B_i, B_j) \\
&= \frac{M\sigma^2}{M^2} + \frac{2}{M^2}\rho\sigma^2\binom{M}{2} \\
&= \underbrace{\rho\sigma^2}_{\text{Resolve with independence}} + \underbrace{\frac{1-\rho}{M}\sigma^2}_{\text{Resolve with }\uparrow\text{classifiers}}
\end{aligned}
$$

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$
- For each bootstrap step m, $(m = 1, 2, \ldots, M)$ draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m$, $\tilde{\boldsymbol{X}}_m$.

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$
- For each bootstrap step m, $(m = 1, 2, \ldots, M)$ draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m$, $\tilde{\boldsymbol{X}}_m$.
- Train classifier on bootstrapped data,

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$
- For each bootstrap step m, $(m = 1, 2, \ldots, M)$ draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m$, $\tilde{\boldsymbol{X}}_m$.
- Train classifier on bootstrapped data,

$$\tilde{\boldsymbol{Y}}_m = f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$
- For each bootstrap step m, $(m = 1, 2, \ldots, M)$ draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m$, $\tilde{\boldsymbol{X}}_m$.
- Train classifier on bootstrapped data,

$$\tilde{\boldsymbol{Y}}_m = f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

- Aggregating across classifiers,

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$
- For each bootstrap step m, $(m = 1, 2, \ldots, M)$ draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m$, $\tilde{\boldsymbol{X}}_m$.
- Train classifier on bootstrapped data,

$$\tilde{\boldsymbol{Y}}_m = f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

- Aggregating across classifiers,

$$f_{\text{bag}}(\boldsymbol{x}_i) = \frac{1}{M}\sum_{m=1}^{M} f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$
- For each bootstrap step m, $(m = 1, 2, \ldots, M)$ draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m$, $\tilde{\boldsymbol{X}}_m$.
- Train classifier on bootstrapped data,

$$\tilde{\boldsymbol{Y}}_m = f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

- Aggregating across classifiers,

$$f_{\text{bag}}(\boldsymbol{x}_i) = \frac{1}{M} \sum_{m=1}^{M} f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

- Only leads to a difference in estimate if classifiers are non-linear.

Bagging: bootstrap aggregation

Creating Weak Classifiers with resampling:

- Suppose we have labels $\boldsymbol{Y}$ and document term matrix $\boldsymbol{X}$
- For each bootstrap step m, $(m = 1, 2, \ldots, M)$ draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m$, $\tilde{\boldsymbol{X}}_m$.
- Train classifier on bootstrapped data,

$$\tilde{\boldsymbol{Y}}_m = f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

- Aggregating across classifiers,

$$f_{\text{bag}}(\boldsymbol{x}_i) = \frac{1}{M}\sum_{m=1}^{M} f^m(\tilde{\boldsymbol{X}}_m, \widehat{\boldsymbol{\beta}}, \boldsymbol{\lambda})$$

- Only leads to a difference in estimate if classifiers are non-linear.
- Strong Correlation between classifiers

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify $\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify$\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

- Within each strata p, compute average Y

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify $\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

- Within each strata p, compute average Y

$$\bar{Y}|\boldsymbol{x}_p = \sum_{i=1}^{N} \frac{I(\boldsymbol{x}_i = \boldsymbol{x}_p) Y_i}{\sum_{t=1}^{N} I(\boldsymbol{x}_t = \boldsymbol{x}_p)}$$

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify$\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

- Within each strata p, compute average Y

$$\bar{Y}|\boldsymbol{x}_p = \sum_{i=1}^{N} \frac{I(\boldsymbol{x}_i = \boldsymbol{x}_p)Y_i}{\sum_{t=1}^{N} I(\boldsymbol{x}_t = \boldsymbol{x}_p)}$$

Implies that for test data we would fit:

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify $\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

- Within each strata p, compute average Y

$$\bar{Y}|\boldsymbol{x}_p = \sum_{i=1}^{N} \frac{I(\boldsymbol{x}_i = \boldsymbol{x}_p)Y_i}{\sum_{t=1}^{N} I(\boldsymbol{x}_t = \boldsymbol{x}_p)}$$

Implies that for test data we would fit:

$$\hat{f}(\boldsymbol{x}_i) = \sum_{p=1}^{P} \bar{Y}|\boldsymbol{x}_p I(\boldsymbol{x}_i = \boldsymbol{x}_p)$$

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify $\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

- Within each strata p, compute average Y

$$\bar{Y}|\boldsymbol{x}_p = \sum_{i=1}^{N} \frac{I(\boldsymbol{x}_i = \boldsymbol{x}_p) Y_i}{\sum_{t=1}^{N} I(\boldsymbol{x}_t = \boldsymbol{x}_p)}$$

Implies that for test data we would fit:

$$\begin{aligned}
\hat{f}(\boldsymbol{x}_i) &= \sum_{p=1}^{P} \bar{Y}|\boldsymbol{x}_p I(\boldsymbol{x}_i = \boldsymbol{x}_p) \\
&= \sum_{p=1}^{P} c_p I(\boldsymbol{x}_i = \boldsymbol{x}_p)
\end{aligned}$$

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify$\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

- Within each strata p, compute average Y

$$\bar{Y}|\boldsymbol{x}_p = \sum_{i=1}^{N} \frac{I(\boldsymbol{x}_i = \boldsymbol{x}_p)Y_i}{\sum_{t=1}^{N} I(\boldsymbol{x}_t = \boldsymbol{x}_p)}$$

Implies that for test data we would fit:

$$\begin{aligned}\hat{f}(\boldsymbol{x}_i) &= \sum_{p=1}^{P} \bar{Y}|\boldsymbol{x}_p I(\boldsymbol{x}_i = \boldsymbol{x}_p) \\ &= \sum_{p=1}^{P} c_p I(\boldsymbol{x}_i = \boldsymbol{x}_p)\end{aligned}$$

Curse of dimensionality(!!!)

Classification and Regression Trees (CART): Intuition

Consider regression $E[Y|\boldsymbol{x}_i]$.

With no assumptions, stratify $\rightsquigarrow$ different mean for unique values of $\boldsymbol{x}_i$

- Within each strata p, compute average Y

$$\bar{Y}|\boldsymbol{x}_p = \sum_{i=1}^{N} \frac{I(\boldsymbol{x}_i = \boldsymbol{x}_p)Y_i}{\sum_{t=1}^{N} I(\boldsymbol{x}_t = \boldsymbol{x}_p)}$$

Implies that for test data we would fit:

$$\begin{aligned}
\hat{f}(\boldsymbol{x}_i) &= \sum_{p=1}^{P} \bar{Y}|\boldsymbol{x}_p I(\boldsymbol{x}_i = \boldsymbol{x}_p) \\
&= \sum_{p=1}^{P} c_p I(\boldsymbol{x}_i = \boldsymbol{x}_p)
\end{aligned}$$

Curse of dimensionality(!!!)

Approximate with regions $\rightsquigarrow$ search for splits of data to approximate stratification

Classification and Regression Trees (CART): Objective function

Labels $\boldsymbol{Y}_i$ and documents $\boldsymbol{x}_i$

$$
\begin{aligned}
E[Y|\boldsymbol{x}_i] &= \widehat{f}(\boldsymbol{x}_i) \\
&= \sum_{p=1}^{P} c_p I(\boldsymbol{x}_i \in R_p)
\end{aligned}
$$

where:

- R_p describes a region $\rightsquigarrow$ node
- c_p describes values of Y_i for document in R_p

Classification and Regression Trees (CART): Optimization function

Suppose we want to minimize sum of squared residuals with each node

Classification and Regression Trees (CART): Optimization function

Suppose we want to minimize sum of squared residuals with each node
Then c_p = Average Y for documents assigned to R_p

Classification and Regression Trees (CART): Optimization function

Suppose we want to minimize sum of squared residuals with each node
Then c_p = Average Y for documents assigned to R_p

$$\widehat{c}_p = \sum_{i=1}^{N} \frac{Y_i I(\boldsymbol{x}_i \in R_p)}{\sum_{j=1}^{N} I(\boldsymbol{x}_j \in R_p)}$$

Classification and Regression Trees (CART): Optimization function

Suppose we want to minimize sum of squared residuals with each node
Then c_p = Average Y for documents assigned to R_p

$$\widehat{c}_p = \sum_{i=1}^{N} \frac{Y_i I(\boldsymbol{x}_i \in R_p)}{\sum_{j=1}^{N} I(\boldsymbol{x}_j \in R_p)}$$

Determining an optimal partition $\rightsquigarrow$ NP-Hard.

Classification and Regression Trees (CART): Optimization function

Suppose we want to minimize sum of squared residuals with each node
Then c_p = Average Y for documents assigned to R_p

$$\widehat{c}_p = \sum_{i=1}^{N} \frac{Y_i I(\boldsymbol{x}_i \in R_p)}{\sum_{j=1}^{N} I(\boldsymbol{x}_j \in R_p)}$$

Determining an optimal partition $\rightsquigarrow$ NP-Hard.
Suppose we are in some node (perhaps at the start).

Classification and Regression Trees (CART): Optimization function

Suppose we want to minimize sum of squared residuals with each node
Then c_p = Average Y for documents assigned to R_p

$$\widehat{c}_p = \sum_{i=1}^{N} \frac{Y_i I(\boldsymbol{x}_i \in R_p)}{\sum_{j=1}^{N} I(\boldsymbol{x}_j \in R_p)}$$

Determining an optimal partition $\rightsquigarrow$ NP-Hard.
Suppose we are in some node (perhaps at the start).
Greedy algorithm:

Classification and Regression Trees (CART): Optimization function

Suppose we want to minimize sum of squared residuals with each node
Then c_p = Average Y for documents assigned to R_p

$$\widehat{c}_p = \sum_{i=1}^{N} \frac{Y_i I(\boldsymbol{x}_i \in R_p)}{\sum_{j=1}^{N} I(\boldsymbol{x}_j \in R_p)}$$

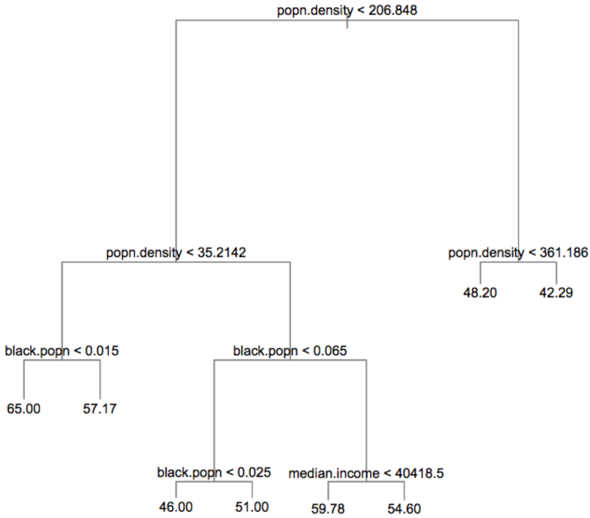
Determining an optimal partition $\rightsquigarrow$ NP-Hard.
Suppose we are in some node (perhaps at the start).
Greedy algorithm:

$$(j^*, s^*) = \arg\min_{j,s} \left[\underbrace{\min_{c_1} \sum_{i=1}^{N} I(x_{ij} < s)(Y_i - c_1)^2}_{\text{“cost” group 1}} + \underbrace{\min_{c_2} \sum_{i=1}^{N} I(x_{ij} > s)(Y_i - c_2)^2}_{\text{“cost” group 2}} \right]$$

Classification and Regression Trees (CART): Algorithm

- Start in Node
- Partition according to Greedy algorithm
- Continue until some stopping rule: number of observations per node

CART Picture (Spirling 2008)



Forests and Trees

Recall: accurate (unbiased) and uncorrelated classifiers

Forests and Trees

Recall: accurate (unbiased) and uncorrelated classifiers

- Grow trees deeply$\rightsquigarrow$ unbiased classifers, though high variance

Forests and Trees

Recall: accurate (unbiased) and uncorrelated classifiers

- Grow trees deeply $\rightsquigarrow$ unbiased classifers, though high variance
- Average $\rightsquigarrow$ reduces variance, but will be correlated

Forests and Trees

Recall: accurate (unbiased) and uncorrelated classifiers

- Grow trees deeply $\rightsquigarrow$ unbiased classifers, though high variance
- Average $\rightsquigarrow$ reduces variance, but will be correlated
- Random forest $\rightsquigarrow$ introduce additional sampling to induce independence $\rightsquigarrow$ Only split on subset of variables

Random Forest Algorithm (ESL, 588)

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:
    i) Select z of the J variables $\rightsquigarrow$ introduces independences across the trees

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:
   i) Select z of the J variables$\rightsquigarrow$ introduces independences across the trees
   ii) Among those z, select the best split node

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:
   i) Select z of the J variables $\rightsquigarrow$ introduces independences across the trees
   ii) Among those z, select the best split node
   iii) Split into daughter nodes

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:
   i) Select z of the J variables $\rightsquigarrow$ introduces independences across the trees
   ii) Among those z, select the best split node
   iii) Split into daughter nodes
3) The result is an ensemble (forest) of trees $\boldsymbol{T} = (T_1, T_2, \ldots, T_M)$,

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:
   i) Select z of the J variables $\rightsquigarrow$ introduces independences across the trees
   ii) Among those z, select the best split node
   iii) Split into daughter nodes
3) The result is an ensemble (forest) of trees $\boldsymbol{T} = (T_1, T_2, \ldots, T_M)$,

$$\hat{f}(\boldsymbol{x}_i) = \frac{1}{M} \sum_{m=1}^{M} T_m(\boldsymbol{x}_i)$$

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:
   i) Select z of the J variables$\rightsquigarrow$ introduces independences across the trees
   ii) Among those z, select the best split node
   iii) Split into daughter nodes
3) The result is an ensemble (forest) of trees $\boldsymbol{T} = (T_1, T_2, \ldots, T_M)$,

$$\hat{f}(\boldsymbol{x}_i) \quad = \quad \frac{1}{M} \sum_{m=1}^{M} T_m(\boldsymbol{x}_i)$$

`RandomForest`$\rightsquigarrow$ Not a silver bullet!

Random Forest Algorithm (ESL, 588)

1) For m bootstrap samples ($m = 1, \ldots, M$), draw N observations with replacement, $\tilde{\boldsymbol{Y}}_m, \tilde{\boldsymbol{X}}_m$
2) Until a minimum node size is reached:
    i) Select z of the J variables$\rightsquigarrow$ introduces independences across the trees
    ii) Among those z, select the best split node
    iii) Split into daughter nodes
3) The result is an ensemble (forest) of trees $\boldsymbol{T} = (T_1, T_2, \ldots, T_M)$,

$$\hat{f}(\boldsymbol{x}_i) = \frac{1}{M}\sum_{m=1}^{M} T_m(\boldsymbol{x}_i)$$

`RandomForest`$\rightsquigarrow$ Not a silver bullet!

- With many poor predictors$\rightsquigarrow$ the p selected may be meaningless

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$
   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$

Super Learning

1) Set of hand labeled documents. For each i, ($i = 1, \ldots, N_{\text{train}}$)

   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$

2) Estimate relationship between labels and words

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$

   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$

2) Estimate relationship between labels and words
   - Each document i is a count vector of K words

Super Learning

1) Set of hand labeled documents. For each i, ($i = 1, \ldots, N_{\text{train}}$)

   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$

2) Estimate relationship between labels and words
   - Each document i is a count vector of K words

     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

Super Learning

1) Set of hand labeled documents. For each i, ($i = 1, \ldots, N_{\text{train}}$)
   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k|\mathbf{x}_i)$$

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$
   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k | \mathbf{x}_i)_{\text{train}}$$

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$

$Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$

2) Estimate relationship between labels and words
   - Each document i is a count vector of K words

     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k | \mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels

Super Learning

1) Set of hand labeled documents. For each i, ($i = 1, \ldots, N_{\text{train}}$)
   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k|\mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels $\rightsquigarrow$ Data and assumptions

Super Learning

1) Set of hand labeled documents. For each i, ($i = 1, \ldots, N_{\text{train}}$)
   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k|\mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels $\rightsquigarrow$ Data and assumptions
  - LASSO (Tibshirani 1996): sparsity

Super Learning

1) Set of hand labeled documents. For each i, ($i = 1, \ldots, N_{\text{train}}$)

   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$

2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k | \mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels $\rightsquigarrow$ Data and assumptions
  - LASSO (Tibshirani 1996): sparsity
  - KRLS (Hainmueller and Hazlett 2013): dense, flexible surface

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$
$Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k | \mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels $\rightsquigarrow$ Data and assumptions
  - LASSO (Tibshirani 1996): sparsity
  - KRLS (Hainmueller and Hazlett 2013): dense, flexible surface
  - Ridge, Elastic-Net, SVM, Random Forests, BART, ...

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$
   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k | \mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels $\rightsquigarrow$ Data and assumptions
  - LASSO (Tibshirani 1996): sparsity
  - KRLS (Hainmueller and Hazlett 2013): dense, flexible surface
  - Ridge, Elastic-Net, SVM, Random Forests, BART, ...
- Which model? Difficult to know before hand

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$
$Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k|\mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels $\rightsquigarrow$ Data and assumptions
  - LASSO (Tibshirani 1996): sparsity
  - KRLS (Hainmueller and Hazlett 2013): dense, flexible surface
  - Ridge, Elastic-Net, SVM, Random Forests, BART, ...
- Which model? Difficult to know before hand
- Assess out of sample performance with cross validation

Super Learning

1) Set of hand labeled documents. For each i, $(i = 1, \ldots, N_{\text{train}})$
   $Y_{i,\text{train}} \in \{C_1 C_2, \ldots C_K\}$
2) Estimate relationship between labels and words
   - Each document i is a count vector of K words
     $\mathbf{x}_{i,\text{train}} = (X_{i1}, X_{i2}, \ldots, X_{iK})$

$$\Pr(Y_i = C_k | \mathbf{x}_i)_{\text{train}} \quad = \quad \widehat{g}(\mathbf{x}_i)_{\text{train}}$$

- Identify systematic relationship between words, labels $\rightsquigarrow$ Data and assumptions
  - LASSO (Tibshirani 1996): sparsity
  - KRLS (Hainmueller and Hazlett 2013): dense, flexible surface
  - Ridge, Elastic-Net, SVM, Random Forests, BART, ...
- Which model? Difficult to know before hand
- Assess out of sample performance with cross validation

Weighted ensemble: weights determined by (unique) out of sample predictive performance

Committee Methods:
Fit many methods, average with equal weights

- Voting (classification)
- Averaging (predictions)

Problem: many poor methods may overwhelm high quality fit (remember earlier figures)
Solution: learn weights via cross validation

Weighted Ensemble to Classify Documents

- Suppose we have M ($m = 1, \ldots, M$) models.

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \mathrm{C}_1 | \mathbf{x}_i)_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

Weighted Ensemble to Classify Documents

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\Pr(Y_i = C_1|\mathbf{x}_i)_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

Weighted Ensemble to Classify Documents

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\Pr(Y_i = C_1 | \mathbf{x}_i)_{\text{train}} = \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \mathrm{C}_1|\mathbf{x}_i)_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights $(\widehat{\pi}_m)$

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \text{C}_1 | \mathbf{x}_i)_{\text{train}} = \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights $(\widehat{\pi}_m)$
  - K-fold cross validation: generate M out of sample predictions for each document in training set

Weighted Ensemble to Classify Documents

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\Pr(Y_i = \mathrm{C}_1 | \mathbf{x}_i)_{\text{train}} = \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights ($\widehat{\pi}_m$)
  - K-fold cross validation: generate M out of sample predictions for each document in training set
    $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$

Weighted Ensemble to Classify Documents

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\Pr(Y_i = \mathrm{C}_1 | \mathbf{x}_i)_{\text{train}} = \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights ($\widehat{\pi}_m$)
  - K-fold cross validation: generate M out of sample predictions for each document in training set $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
  - Estimate weights with constrained regression:

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \mathrm{C}_1|\mathbf{x}_i)_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights $(\widehat{\pi}_m)$
  - K-fold cross validation: generate M out of sample predictions for each document in training set $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
  - Estimate weights with constrained regression:

$$Y_i \quad = \quad \sum_{m=1}^{M} \pi_m \hat{Y}_{im} + \epsilon_i$$

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \text{C}_1 | \mathbf{x}_i)_{\text{train}} = \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights $(\widehat{\pi}_m)$
    - K-fold cross validation: generate M out of sample predictions for each document in training set
      $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
    - Estimate weights with constrained regression:

$$Y_i = \sum_{m=1}^{M} \pi_m \hat{Y}_{im} + \epsilon_i$$

where we impose constraints: $\pi_m \geq 0$ and $\sum_{m=1}^{M} \pi_m = 1$.

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \mathrm{C}_1 | \mathbf{x}_i)_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights $(\widehat{\pi}_m)$
    - K-fold cross validation: generate M out of sample predictions for each document in training set $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
    - Estimate weights with constrained regression:

$$Y_i \quad = \quad \sum_{m=1}^{M} \pi_m \hat{Y}_{im} + \epsilon_i$$

where we impose constraints: $\pi_m \geq 0$ and $\sum_{m=1}^{M} \pi_m = 1$.

- Result $\widehat{\pi}_m$ for each method

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \text{C}_1 | \mathbf{x}_i)_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights $(\widehat{\pi}_m)$
- Estimate $\widehat{g}_m(\mathbf{x}_i) \rightsquigarrow$ Apply all M models to entire training set

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \mathrm{C}_1 | \mathbf{x}_i)_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_i)$$

- Estimate weights $(\widehat{\pi}_m)$
- Estimate $\widehat{g}_m(\mathbf{x}_i) \rightsquigarrow$ Apply all M models to entire training set

3) For each document i in test set, $\mathbf{x}_{i,\text{test}}$

Weighted Ensemble to Classify Documents

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\Pr(Y_i = C_1 | \mathbf{x}_{i,\text{test}})_{\text{train}} = \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_{i,\text{test}})$$

- Estimate weights ($\widehat{\pi}_m$)
- Estimate $\widehat{g}_m(\mathbf{x}_i) \rightsquigarrow$ Apply all M models to entire training set

3) For each document i in test set, $\mathbf{x}_{i,\text{test}}$

Weighted Ensemble to Classify Documents

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\Pr(Y_i = \mathrm{C}_1 | \mathbf{x}_{i,\text{test}})_{\text{train}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m \widehat{g}_m(\mathbf{x}_{i,\text{test}})$$

- Estimate weights $(\widehat{\pi}_m)$
- Estimate $\widehat{g}_m(\mathbf{x}_i) \rightsquigarrow$ Apply all M models to entire training set

3) For each document i in test set, $\mathbf{x}_{i,\text{test}}$
(Classify if above threshold)

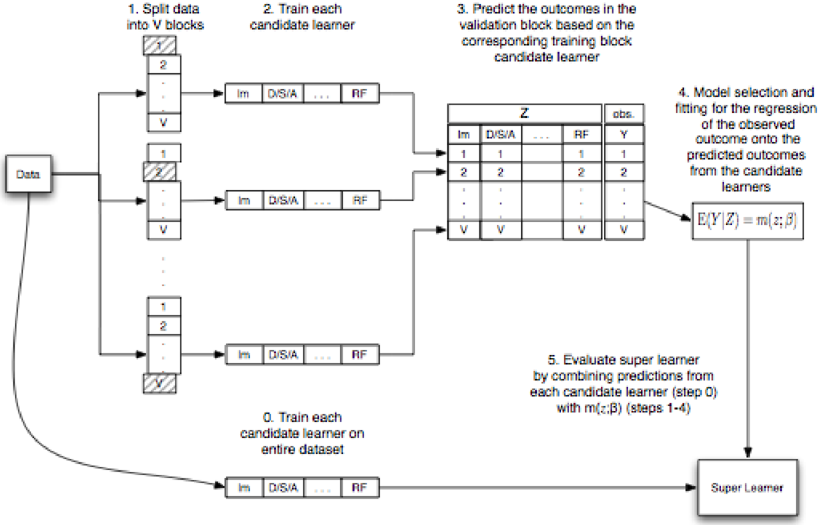


Figure 1: Flow Diagram for Super Learner

Why Super Learn?

van der Laan et al (2007) prove:

- Asymptotically: super learners will perform as well the best candidates for data
- Oracle: performs like the best possible method among candidate methods
  - Asymptotically outperforms constituent methods
  - Performs as well as optimal combinations of those methods

Practical questions:

- Final regression:
  - Logistic
  - Linear
  - Could super learn again!
- How Many Folds?
  - van der Laan et al's proofs rely on growing folds with N (but slowly)
  - Use 10-fold cross validation for simulations

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting $\rightsquigarrow$ few disagreements

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting$\rightsquigarrow$ few disagreements

Use five classifiers to form Ensemble (cross validating within each to tune parameters)

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting$\rightsquigarrow$ few disagreements

Use five classifiers to form Ensemble (cross validating within each to tune parameters)

- LASSO

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting$\rightsquigarrow$ few disagreements

Use five classifiers to form Ensemble (cross validating within each to tune parameters)

- LASSO
- Elastic-Net

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting$\rightsquigarrow$ few disagreements

Use five classifiers to form Ensemble (cross validating within each to tune parameters)

- LASSO
- Elastic-Net
- Random Forest

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting $\rightsquigarrow$ few disagreements

Use five classifiers to form Ensemble (cross validating within each to tune parameters)

- LASSO
- Elastic-Net
- Random Forest
- A Support Vector Machine

Impression of Influence

Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting $\rightsquigarrow$ few disagreements

Use five classifiers to form Ensemble (cross validating within each to tune parameters)

- LASSO
- Elastic-Net
- Random Forest
- A Support Vector Machine
- Kernel Regularized Least Squares (KRLS, Hainmueller and Hazlett 2014)

Impression of Influence

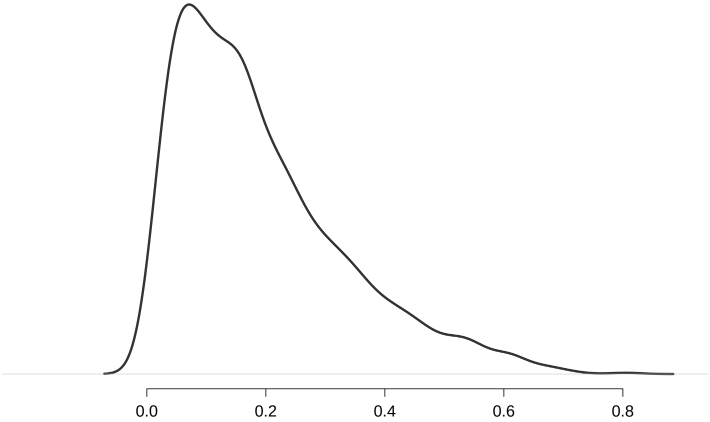
Estimate: $Y_i \in \{\text{Credit}, \text{Not Credit}\}$

- Triple hand code 800 press releases
- Resolve disagreement with voting$\leadsto$ few disagreements

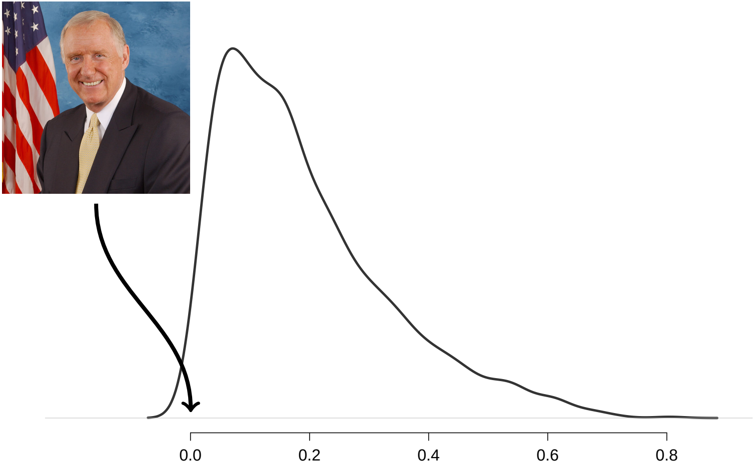
Use five classifiers to form Ensemble (cross validating within each to tune parameters)

- LASSO 0
- Elastic-Net 23%
- Random Forest 61%
- A Support Vector Machine16%
- Kernel Regularized Least Squares (KRLS, Hainmueller and Hazlett 2014) 0

Strategic Credit Claiming to Build a Personal Vote

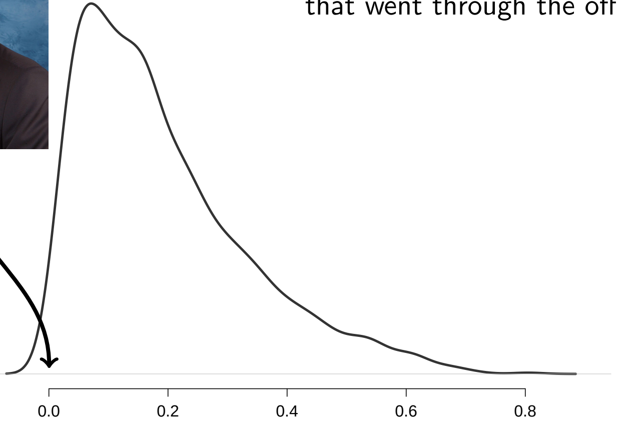


Strategic Credit Claiming to Build a Personal Vote



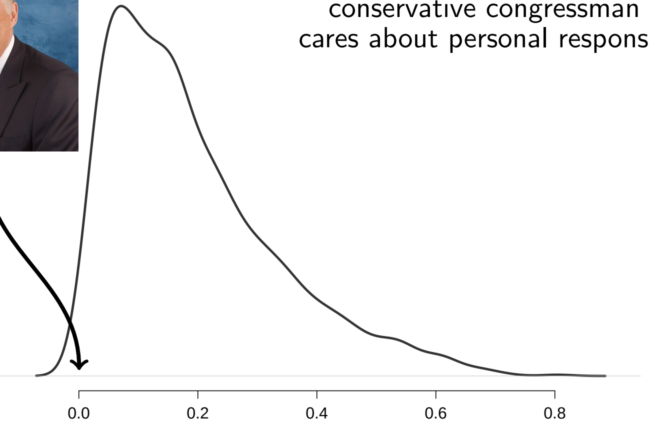
Strategic Credit Claiming to Build a Personal Vote

John McGroff: "voted for every spending bill " that went through the office"

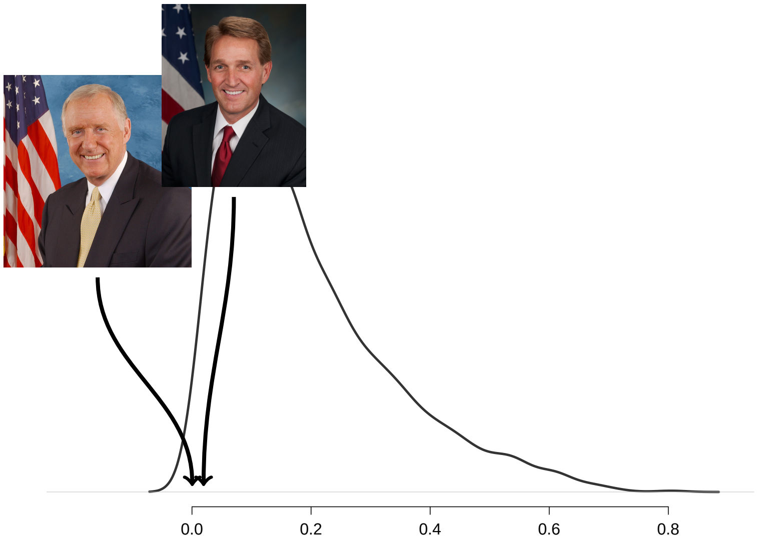


Strategic Credit Claiming to Build a Personal Vote

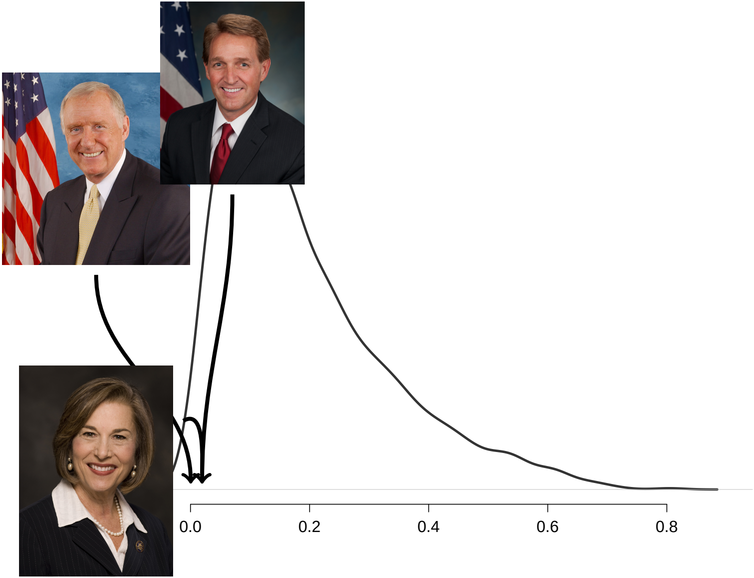
John McGroff: “Not the actions of a fiscally ”
conservative congressman who
cares about personal responsibility”



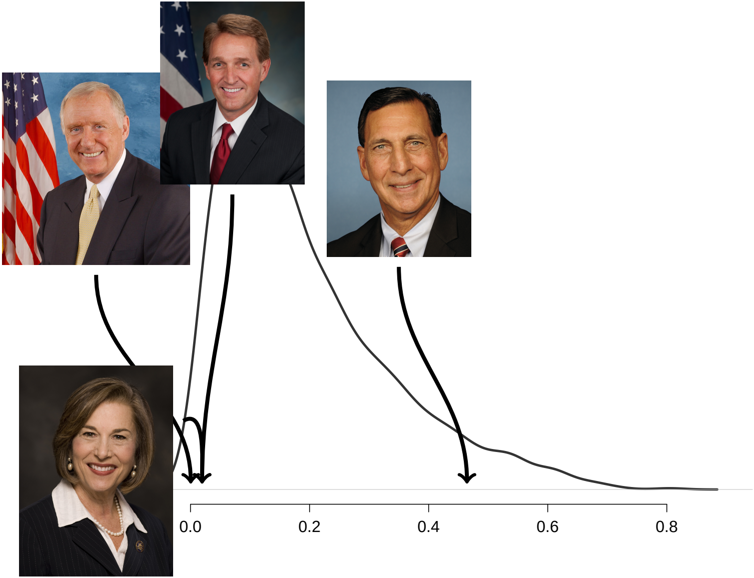
Strategic Credit Claiming to Build a Personal Vote



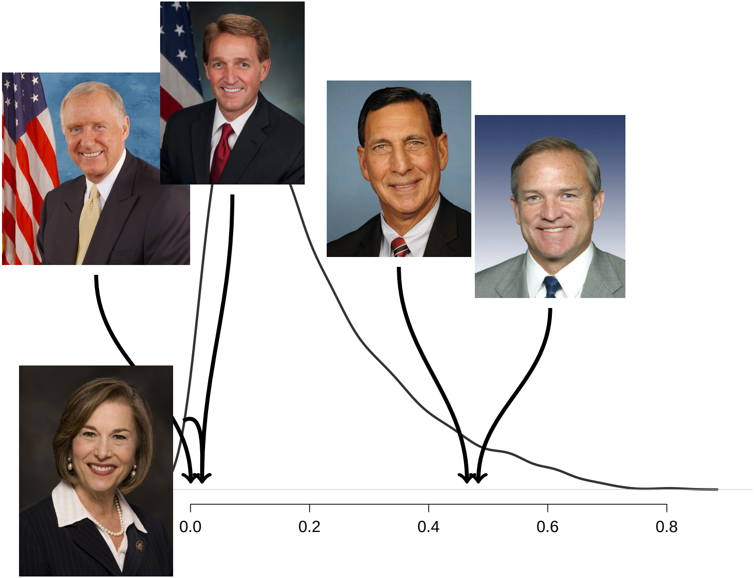
Strategic Credit Claiming to Build a Personal Vote



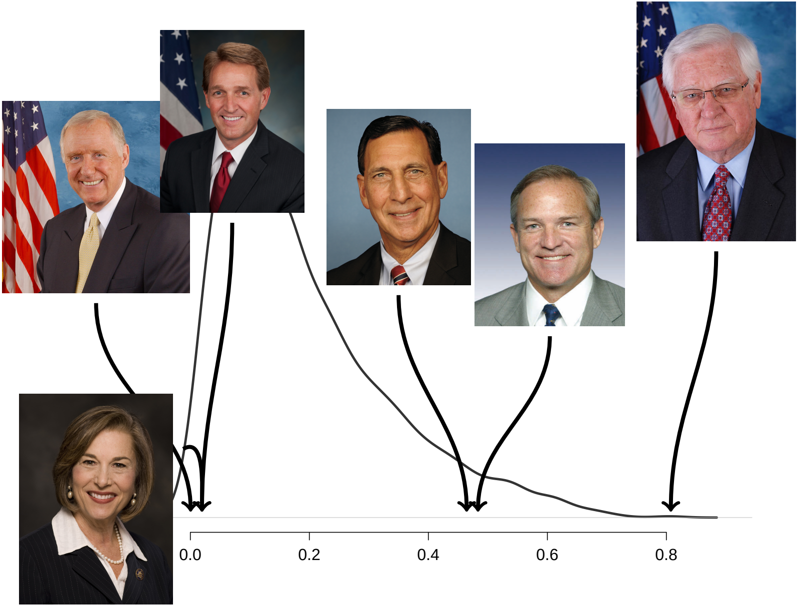
Strategic Credit Claiming to Build a Personal Vote



Strategic Credit Claiming to Build a Personal Vote

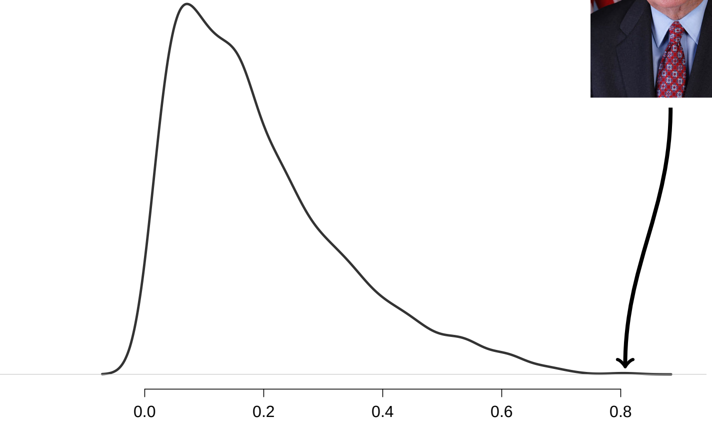


Strategic Credit Claiming to Build a Personal Vote



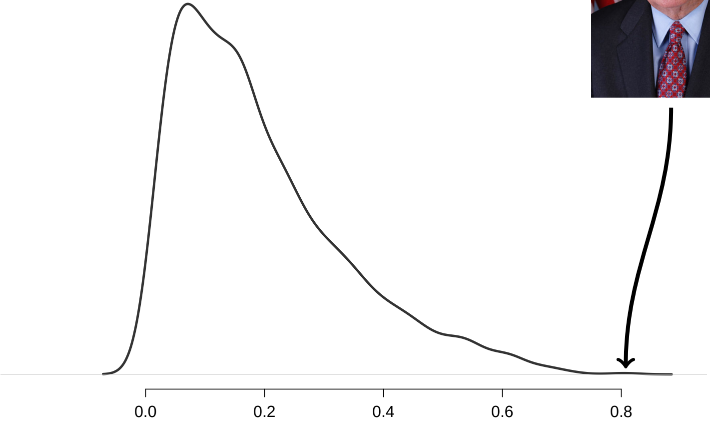
Strategic Credit Claiming to Build a Personal Vote

“We just can’t afford luxuries like ideology”



Strategic Credit Claiming to Build a Personal Vote

Lexington Herald-Leader: Prince of Pork



Other Reasons to Ensemble (Dietterich 2000)

Statistical

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

- Methods stuck in local modes

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

- Methods stuck in local modes
- Result: no one run provides best model

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

- Methods stuck in local modes
- Result: no one run provides best model
- Averages of runs may perform better

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

- Methods stuck in local modes
- Result: no one run provides best model
- Averages of runs may perform better

Complex "true" functional forms

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

- Methods stuck in local modes
- Result: no one run provides best model
- Averages of runs may perform better

Complex "true" functional forms

- One method may be unable to approximate true DGP

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

- Methods stuck in local modes
- Result: no one run provides best model
- Averages of runs may perform better

Complex "true" functional forms

- One method may be unable to approximate true DGP
- Mixtures of methods may approximate better

Other Reasons to Ensemble (Dietterich 2000)

Statistical

- With little data, many algorithms offer similar performance
- Ensemble ensures we avoid wrong model in test set

Computational

- Methods stuck in local modes
- Result: no one run provides best model
- Averages of runs may perform better

Complex "true" functional forms

- One method may be unable to approximate true DGP
- Mixtures of methods may approximate better

Ensembles Beyond Text Data

Machine Learning $\leftrightarrow$ Causal Inference

An Example Experiment

An Example Experiment

Rep. Harold “Hal” Rogers (KY-05) announced today that Kentucky is slated to receive $962,500 to protect critical infrastructure- power plants, chemical facilities, stadiums, and other high-risk assets, through the U.S. Department of Homeland Security’s buffer zone protection program

An Example Experiment

A federal grant will help keep the Brainerd Lakes Airport operating in winter weather. Today, Congressman Jim Oberstar announced that the Federal Aviation Administration (FAA) will award $528,873 to the Brainerd airport. The funding will be used to purchase new snow removal and deicing equipment.

An Example Experiment

Congresswoman Darlene Hooley (OR-5) and Congressmen Earl Blumenauer (OR-3), David Wu (OR-1) and Greg Walden (OR-2) joined together today in announcing $375,000 in federal funding for the Oregon Partnership to combat methamphetamine abuse in Oregon.

An Example Experiment

What information in credit claiming messages affect evaluations?

Rewarding Actions and Type of Expenditure, Not Money

Experiment: vary the recipient of money and the action reported in credit claiming statement (and many other features)

Rewarding Actions and Type of Expenditure, Not Money

Experiment: vary the recipient of money and the action reported in credit claiming statement (and many other features)
Treatments: type

1) Planned Parenthood
2) Parks
3) Gun Range
4) Fire Department
5) Police
6) Roads

Rewarding Actions and Type of Expenditure, Not Money

Experiment: vary the recipient of money and the action reported in credit claiming statement (and many other features)
Treatments: type, stage

1) Will request
2) Requested
3) Secured

Rewarding Actions and Type of Expenditure, Not Money

Experiment: vary the recipient of money and the action reported in credit claiming statement (and many other features)
Treatments: type, stage, money

1) \$50 Thousand
2) \$20 Million

Rewarding Actions and Type of Expenditure, Not Money

Experiment: vary the recipient of money and the action reported in credit claiming statement (and many other features)
Treatments: type, stage, money, collaboration

1) Alone
2) w/ Senate Democrat
3) w/ Senate Republican

Rewarding Actions and Type of Expenditure, Not Money

Experiment: vary the recipient of money and the action reported in credit claiming statement (and many other features)
Treatments: type, stage, money, collaboration, partisanship

1) Democrat
2) Republican

Rewarding Actions and Type of Expenditure, Not Money

Experiment: vary the recipient of money and the action reported in credit claiming statement (and many other features)
Treatments: type, stage, money, collaboration, partisanship
Control Condition:

Advertising press release

Rewarding Actions and Type of Expenditure, Not Money

Example Treatment:

Headline: Representative [blackbox] secured $50 Thousand to purchase safety equipment for local firefighters

Body: Representative [blackbox] (Democrat) and Senator [blackbox], a Democrat, secured $50 Thousand to purchase safety equipment for local firefighters.

Rep. [blackbox] said "This money will help our brave firefighters stay safe as they protect our businesses and homes"

Rewarding Actions and Type of Expenditure, Not Money

Example Treatment:
Headline: Representative [blackbox] will request $20 million for medical equipment at the local Planned Parenthood.
Body: Representative [blackbox] (Democrat), will request $20 million for medical equipment at the local Planned Parenthood.
Rep. [blackbox] said "This money would help provide state of the art care for women in our community."

Rewarding Actions and Type of Expenditure, Not Money

214 other conditions

Rewarding Actions and Type of Expenditure, Not Money

214 other conditions
Dependent variable: Approve of representative

Rewarding Actions and Type of Expenditure, Not Money

214 other conditions

Dependent variable: Approve of representative

Goal $\rightsquigarrow$ measure effect of credit claiming content on approval ratings

Rewarding Actions and Type of Expenditure, Not Money

214 other conditions
Dependent variable: Approve of representative

Goal $\rightsquigarrow$ measure effect of credit claiming content on approval ratings
Mechanics$\rightsquigarrow$ Mechanical Turk sample (Findings are replicated in representative samples, using real representatives/senators)

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Effect of particular component of message:

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Effect of particular component of message:
  - $\mathrm{T}_{\text{stage}} = \text{Secured}$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\text{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\text{T}_{i,\text{type}}, \text{T}_{i,\text{stage}}, \text{T}_{i,\text{money}}, \text{T}_{i,\text{collab.}}, \text{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Effect of particular component of message:
  - $\text{T}_{\text{stage}} = \text{Secured}$
  - $\text{T}_{\text{stage}} = \text{Requested}$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Effect of particular component of message:
  - $\mathrm{T}_{\text{stage}} =$ Secured
  - $\mathrm{T}_{\text{stage}} =$ Requested
  - $\mathrm{T}_{\text{stage}} =$ Will Request

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Effect of particular component of message:
  - $\mathrm{T}_{\text{stage}} = \text{Secured}$
  - $\mathrm{T}_{\text{stage}} = \text{Requested}$
  - $\mathrm{T}_{\text{stage}} = \text{Will Request}$
  - $\mathrm{T}_j = k$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Marginal Average Treatment Effect ($\text{MATE}_{\mathrm{T}_j = k}$)

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Marginal Average Treatment Effect ($\text{MATE}_{\mathrm{T}_j=k}$)

$$\text{MATE}_{\mathrm{T}_j=k} \quad = \quad \int \mathrm{E}[Y(\mathrm{T}_j = k, \mathbf{T}_{-j}) - Y(0)]\mathrm{dF}_{\mathbf{T}_{-j}|\mathrm{T}_j=k}$$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Marginal Average Treatment Effect ($\text{MATE}_{\mathrm{T}_j = k}$)

$$\begin{aligned}
\text{MATE}_{\mathrm{T}_j=k} &= \int \mathrm{E}[Y(\mathrm{T}_j = k, \mathbf{T}_{-j}) - Y(0)] \mathrm{dF}_{\mathbf{T}_{-j}|\mathrm{T}_j=k} \\
\text{MATE}_{\mathrm{T}_j=k} &= \mathrm{E}[Y(\mathrm{T}_j = k) - Y(0)]
\end{aligned}$$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Marginal Average Treatment Effect ($\text{MATE}_{\mathrm{T}_j = k}$)

$$\text{MATE}_{\mathrm{T}_j = k} = \mathrm{E}[Y(\mathrm{T}_j = k)|\mathrm{T}_j = k] - \mathrm{E}[Y(0)|\mathrm{T} = 0]$$

Rewarding Actions and Type of Expenditure, Not Money

- Participant i ($i = 1, \ldots, N$), has treatment assignment $\mathbf{T}_i$
- If $\mathrm{T}_i = 0$ for control condition
- $\mathbf{T}_i = (\mathrm{T}_{i,\text{type}}, \mathrm{T}_{i,\text{stage}}, \mathrm{T}_{i,\text{money}}, \mathrm{T}_{i,\text{collab.}}, \mathrm{T}_{i,\text{part.}})$
- $Y_i(\mathbf{T}_i)$: participant i's Approval decision under treatment $\mathbf{T}_i$
- Quantities of Interest
- Marginal Average Treatment Effect ($\text{MATE}_{\mathrm{T}_j=k}$)

$$\begin{aligned}
\text{MATE}_{\mathrm{T}_j=k} &= \mathrm{E}[Y(\mathrm{T}_j = k)|\mathrm{T}_j = k] - \mathrm{E}[Y(0)|\mathrm{T} = 0] \\
\widehat{\text{MATE}}_{\mathrm{T}_j=k} &= \frac{\sum_{i=1}^{N} Y_i I(\mathrm{T}_{ij} = k)}{\sum_{i=1}^{N} I(\mathrm{T}_{ij} = k)} - \frac{\sum_{i=1}^{N} Y_i I(\mathrm{T}_i = 0)}{\sum_{i=1}^{N} I(\mathrm{T}_i = 0)}
\end{aligned}$$

Rewarding Actions and Type of Expenditure, Not Money

- Response may be conditional on respondent characteristics $\mathbf{x}$

Rewarding Actions and Type of Expenditure, Not Money

- Response may be conditional on respondent characteristics $\mathbf{x}$
- For example $\mathbf{x} =$ (Conservative, Republican)

Rewarding Actions and Type of Expenditure, Not Money

- Response may be conditional on respondent characteristics $\mathbf{x}$
- For example $\mathbf{x} =$ (Conservative, Republican)
- Marginal Conditional Average Treatment Effect ($\text{MCATE}_{\text{T}_j=k,\mathbf{x}}$)

Rewarding Actions and Type of Expenditure, Not Money

- Response may be conditional on respondent characteristics $\mathbf{x}$
- For example $\mathbf{x} =$ (Conservative, Republican)
- Marginal Conditional Average Treatment Effect ($\text{MCATE}_{\text{T}_j=k,\mathbf{x}}$)

$$\text{MCATE}_{\text{T}_j=k,\mathbf{x}} = \text{E}[Y(\text{T}_j = k) - Y(0)|\mathbf{x}]$$

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\text{T}_j = k, \mathbf{x}} \quad = \quad \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$

Rewarding Actions and Type of Expenditure, Not Money

$$
\begin{aligned}
\text{MCATE}_{\mathrm{T}_j=k,\mathbf{x}} &= \mathrm{E}[Y(\mathrm{T}_j = k)|\mathbf{x}] - \mathrm{E}[Y(0)|\mathbf{x}] \\
\widehat{\text{MCATE}}_{\mathrm{T}_j=k,\mathbf{x}} &= \frac{\sum_{i=1}^{N} Y_i I(\mathrm{T}_j = k, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\mathrm{T}_j = k, \mathbf{x}_i = \mathbf{x})} - \frac{\sum_{i=1}^{N} Y_i I(\mathrm{T}_i = 0, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\mathrm{T}_i = 0, \mathbf{x}_i = \mathbf{x})}
\end{aligned}
$$

Rewarding Actions and Type of Expenditure, Not Money

$$
\begin{aligned}
\text{MCATE}_{\text{T}_j = k, \mathbf{x}} &= \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}] \\
\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} &= \frac{\sum_{i=1}^{N} Y_i I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})} - \frac{\sum_{i=1}^{N} Y_i I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}
\end{aligned}
$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\text{T}_j = k, \mathbf{x}} = \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \frac{\sum_{i=1}^{N} Y_i I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})} - \frac{\sum_{i=1}^{N} Y_i I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\text{T}_j = k, \mathbf{x}} = \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \frac{\sum_{i=1}^{N} Y_i I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})} - \frac{\sum_{i=1}^{N} Y_i I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions Heterogeneous treatment effect methods

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\text{T}_j=k,\mathbf{x}} = \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$

$$\widehat{\text{MCATE}}_{\text{T}_j=k,\mathbf{x}} = \frac{\sum_{i=1}^{N} Y_i I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})} - \frac{\sum_{i=1}^{N} Y_i I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions
  Heterogeneous treatment effect methods
  - LASSO, Find It (Imai and Ratkovic, 2013)$\rightsquigarrow$ sparsity

Rewarding Actions and Type of Expenditure, Not Money

$$\mathrm{MCATE}_{\mathrm{T}_j=k,\mathbf{x}} = \mathrm{E}[Y(\mathrm{T}_j = k)|\mathbf{x}] - \mathrm{E}[Y(0)|\mathbf{x}]$$

$$\widehat{\mathrm{MCATE}}_{\mathrm{T}_j=k,\mathbf{x}} = \frac{\sum_{i=1}^N Y_i I(\mathrm{T}_j = k, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^N I(\mathrm{T}_j = k, \mathbf{x}_i = \mathbf{x})} - \frac{\sum_{i=1}^N Y_i I(\mathrm{T}_i = 0, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^N I(\mathrm{T}_i = 0, \mathbf{x}_i = \mathbf{x})}$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions
  Heterogeneous treatment effect methods
  - LASSO, Find It (Imai and Ratkovic, 2013) $\rightsquigarrow$ sparsity
  - Ridge, KRLS (Hainmueller and Hazlett, 2013) $\rightsquigarrow$ flexible surface, dense

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\text{T}_j=k,\mathbf{x}} = \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$

$$\widehat{\text{MCATE}}_{\text{T}_j=k,\mathbf{x}} = \frac{\sum_{i=1}^{N} Y_i I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_j = k, \mathbf{x}_i = \mathbf{x})} - \frac{\sum_{i=1}^{N} Y_i I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}{\sum_{i=1}^{N} I(\text{T}_i = 0, \mathbf{x}_i = \mathbf{x})}$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions
  Heterogeneous treatment effect methods
  - LASSO, Find It (Imai and Ratkovic, 2013)$\rightsquigarrow$ sparsity
  - Ridge, KRLS (Hainmueller and Hazlett, 2013)$\rightsquigarrow$ flexible surface, dense
  - Model m to estimate some function $g_m(\text{T}_j = k, \boldsymbol{x})$

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\mathrm{T}_j=k,\mathbf{x}} = \text{E}[Y(\mathrm{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$
$$\widehat{\text{MCATE}}_{\mathrm{T}_j=k,\mathbf{x}} = \widehat{g}_m(\mathrm{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x})$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions
  Heterogeneous treatment effect methods
  - LASSO, Find It (Imai and Ratkovic, 2013)$\rightsquigarrow$ sparsity
  - Ridge, KRLS (Hainmueller and Hazlett, 2013)$\rightsquigarrow$ flexible surface, dense
  - Model m to estimate some function $g_m(\mathrm{T}_j = k, \boldsymbol{x})$
- Perform well: $g_m(\mathrm{T}_j = k, \boldsymbol{x})$ accurately estimates response surface ($\text{E}[Y(\mathrm{T}_j = k)|\mathbf{x}]$)

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\text{T}_j=k,\mathbf{x}} = \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$

$$\widehat{\text{MCATE}}_{\text{T}_j=k,\mathbf{x}} = \widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x})$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions Heterogeneous treatment effect methods
  - LASSO, Find It (Imai and Ratkovic, 2013)$\rightsquigarrow$ sparsity
  - Ridge, KRLS (Hainmueller and Hazlett, 2013)$\rightsquigarrow$ flexible surface, dense
  - Model m to estimate some function $g_m(\text{T}_j = k, \boldsymbol{x})$
- Perform well: $g_m(\text{T}_j = k, \boldsymbol{x})$ accurately estimates response surface ($\text{E}[Y(\text{T}_j = k)|\mathbf{x}]$)
- Perform well: accurate out of sample prediction and classification (van der Laan et al 2007, Raftery et al 2005)

Rewarding Actions and Type of Expenditure, Not Money

$$\text{MCATE}_{\text{T}_j=k,\mathbf{x}} = \text{E}[Y(\text{T}_j = k)|\mathbf{x}] - \text{E}[Y(0)|\mathbf{x}]$$

$$\widehat{\text{MCATE}}_{\text{T}_j=k,\mathbf{x}} = \widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x})$$

- Curse of Dimensionality: highly variable estimates, (sometimes) empty strata
- Separate systematic differences from noise $\rightsquigarrow$ data and assumptions
  Heterogeneous treatment effect methods
  - LASSO, Find It (Imai and Ratkovic, 2013)$\rightsquigarrow$ sparsity
  - Ridge, KRLS (Hainmueller and Hazlett, 2013)$\rightsquigarrow$ flexible surface, dense
  - Model m to estimate some function $g_m(\text{T}_j = k, \boldsymbol{x})$
- Perform well: $g_m(\text{T}_j = k, \boldsymbol{x})$ accurately estimates response surface $(\text{E}[Y(\text{T}_j = k)|\mathbf{x}])$
- Perform well: accurate out of sample prediction and classification (van der Laan et al 2007, Raftery et al 2005)

Create ensemble: weighting methods by (unique) out of sample predictive performance

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M $(m = 1, \ldots, M)$ models.

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m (\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\text{T}_j=k,\mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m \left(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}) \right)$$

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m (\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights ($\widehat{\pi}_m$)

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\text{T}_j=k,\mathbf{x}} \;=\; \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights ($\widehat{\pi}_m$)
    - 10-fold cross validation: generate M out of sample predictions for each observation

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} \;=\; \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights $(\widehat{\pi}_m)$
    - 10-fold cross validation: generate M out of sample predictions for each observation $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\mathrm{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m (\widehat{g}_m(\mathrm{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights ($\widehat{\pi}_m$)
  - 10-fold cross validation: generate M out of sample predictions for each observation
    $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
  - Estimate weights with constrained regression:

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights ($\widehat{\pi}_m$)
  - 10-fold cross validation: generate M out of sample predictions for each observation
    $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
  - Estimate weights with constrained regression:

$$Y_i = \sum_{m=1}^{M} \pi_m \hat{Y}_{im} + \epsilon_i$$

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\widehat{\text{MCATE}}_{\text{T}_j=k,\mathbf{x}} \quad = \quad \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights $(\widehat{\pi}_m)$
  - 10-fold cross validation: generate M out of sample predictions for each observation
    $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
  - Estimate weights with constrained regression:

$$Y_i \quad = \quad \sum_{m=1}^{M} \pi_m \hat{Y}_{im} + \epsilon_i$$

where we impose constraints: $\pi_m \geq 0$ and $\sum_{m=1}^{M} \pi_m = 1$.

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights ($\widehat{\pi}_m$)
  - 10-fold cross validation: generate M out of sample predictions for each observation
    $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
  - Estimate weights with constrained regression:

$$Y_i = \sum_{m=1}^{M} \pi_m \hat{Y}_{im} + \epsilon_i$$

    where we impose constraints: $\pi_m \geq 0$ and $\sum_{m=1}^{M} \pi_m = 1$.
  - Result $\widehat{\pi}_m$ for each method

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights $(\widehat{\pi}_m)$
  - 10-fold cross validation: generate M out of sample predictions for each observation $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \widehat{Y}_{i2}, \ldots, \widehat{Y}_{iM})$
  - (Alternatively) Estimate weights from mixture model (EBMA) (Raftery et al 2005; Montgomery, Hollenback, Ward 2012) $\rightsquigarrow$ EM, Gibbs, Variational Approximation

Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M $(m = 1, \ldots, M)$ models.

$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}} = \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}) - \widehat{g}_m(0, \mathbf{x}))$$

- Estimate weights $(\widehat{\pi}_m)$
- Estimate $\widehat{g}_m(\text{T}_j = k, \mathbf{x}) \rightsquigarrow$ Apply all M models to entire data set

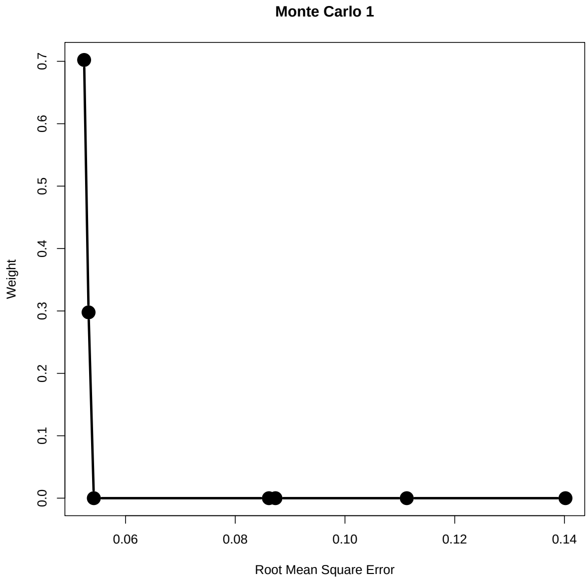
Weighted Ensemble to Measure Credit Claiming Rate

- Suppose we have M ($m = 1, \ldots, M$) models.

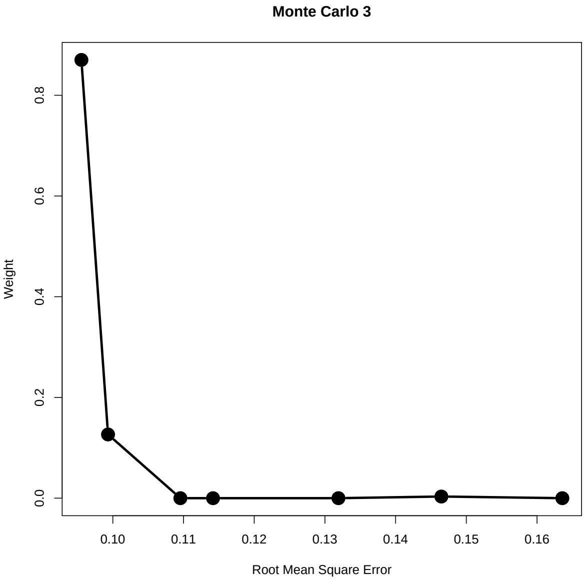
$$\widehat{\text{MCATE}}_{\text{T}_j = k, \mathbf{x}_{\text{new}}} = \sum_{m=1}^{M} \widehat{\pi}_m(\widehat{g}_m(\text{T}_j = k, \mathbf{x}_{\text{new}}) - \widehat{g}_m(0, \mathbf{x}_{\text{new}}))$$

- Estimate weights ($\widehat{\pi}_m$)
- Estimate $\widehat{g}_m(\text{T}_j = k, \mathbf{x})$ $\rightsquigarrow$ Apply all M models to entire data set
- Generate effects of interest (perhaps weighting to other population) $\mathbf{x}_{\text{new}}$

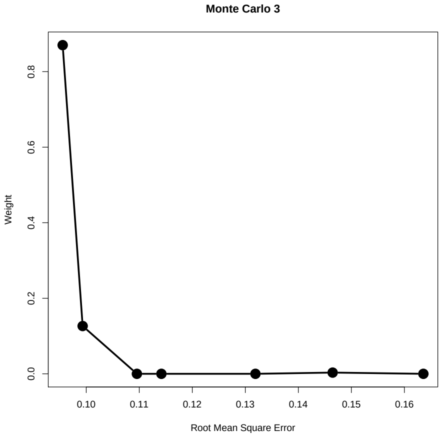
Monte Carlo Evidence



Monte Carlo Evidence



Monte Carlo Evidence



Ensembles outperform constituent methods $\rightsquigarrow$ ensembles place weight on better performing method

Returning to Example Experiment

Recall: experiment to assess effects of credit claiming on approval $\rightsquigarrow$ 1,074 participants (MTurk)

Returning to Example Experiment

Recall: experiment to assess effects of credit claiming on approval $\rightsquigarrow$ 1,074 participants (MTurk)
Apply ensemble method (7 constituent methods, 10 fold cross validation), including treatments and Partisanship and Ideology.

Returning to Example Experiment

Recall: experiment to assess effects of credit claiming on approval $\rightsquigarrow$ 1,074 participants (MTurk)

Apply ensemble method (7 constituent methods, 10 fold cross validation), including treatments and Partisanship and Ideology.

Positive weight on three methods:

Returning to Example Experiment

Recall: experiment to assess effects of credit claiming on approval $\rightsquigarrow$ 1,074 participants (MTurk)
Apply ensemble method (7 constituent methods, 10 fold cross validation), including treatments and Partisanship and Ideology.
Positive weight on three methods:

1) LASSO (0.62)

Returning to Example Experiment

Recall: experiment to assess effects of credit claiming on approval $\rightsquigarrow$ 1,074 participants (MTurk)

Apply ensemble method (7 constituent methods, 10 fold cross validation), including treatments and Partisanship and Ideology.

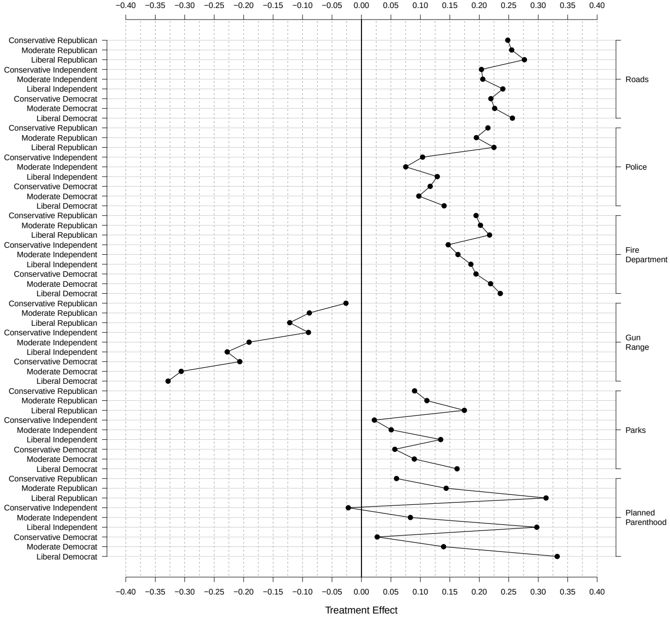
Positive weight on three methods:

1) LASSO (0.62)
2) KRLS (0.24)

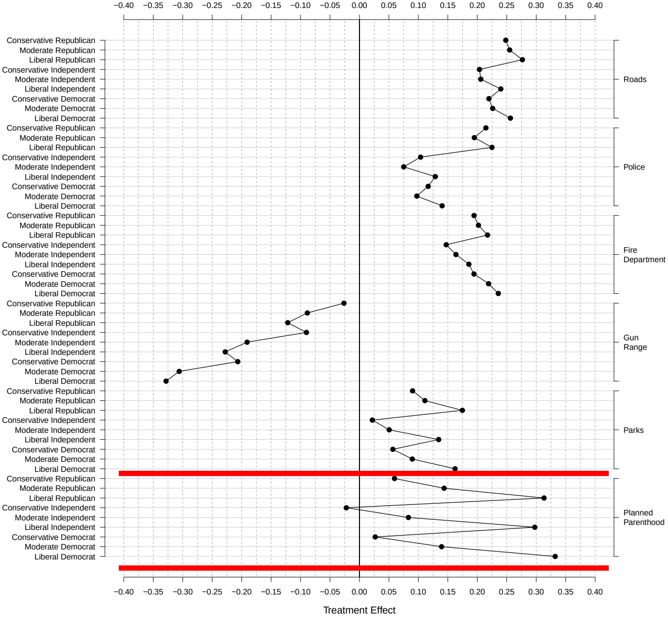
Returning to Example Experiment

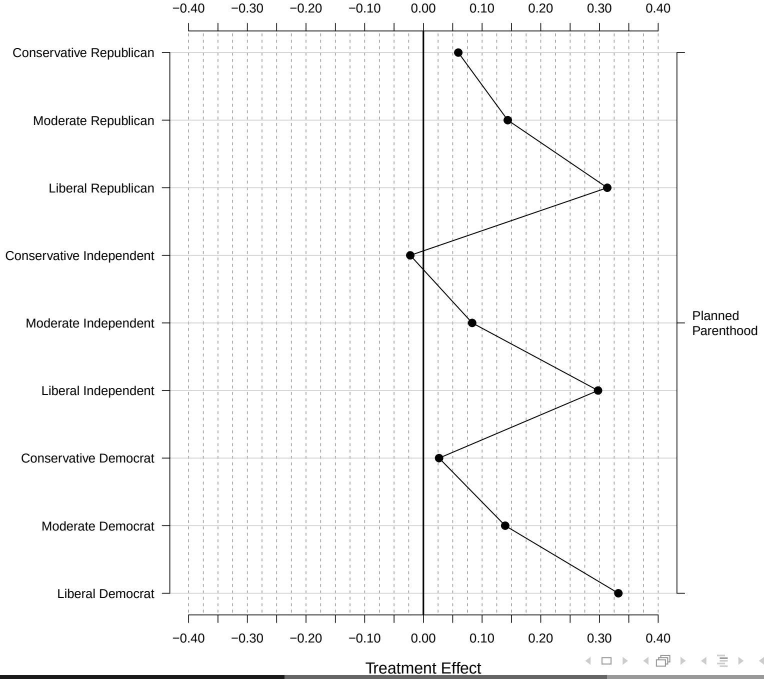
Recall: experiment to assess effects of credit claiming on approval $\rightsquigarrow$ 1,074 participants (MTurk)
Apply ensemble method (7 constituent methods, 10 fold cross validation), including treatments and Partisanship and Ideology.
Positive weight on three methods:

1) LASSO (0.62)
2) KRLS (0.24)
3) Find it (0.14)

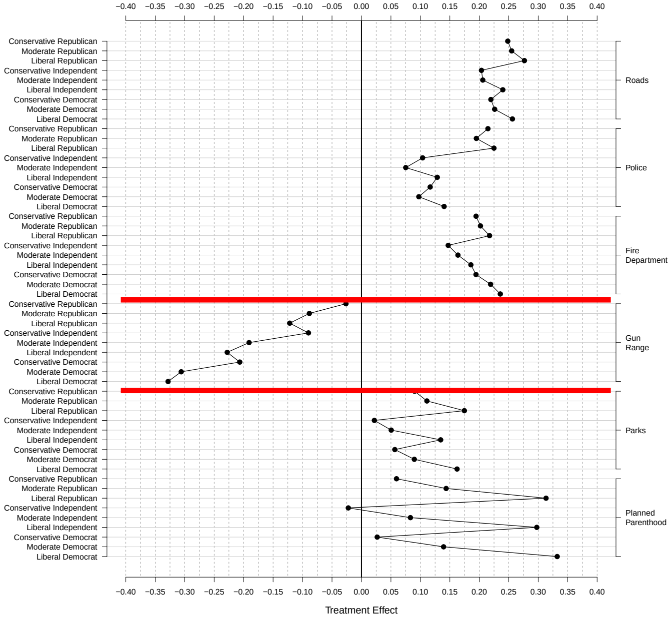


−0.40 −0.35 −0.30 −0.25 −0.20 −0.15 −0.10 −0.05 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
Conservative Republican
Moderate Republican
Liberal Republican
Conservative Independent
Moderate Independent
Liberal Independent
Conservative Democrat
Moderate Democrat
Liberal Democrat
Roads
Police
Fire Department
Gun Range
Parks
Planned Parenthood
Treatment Effect

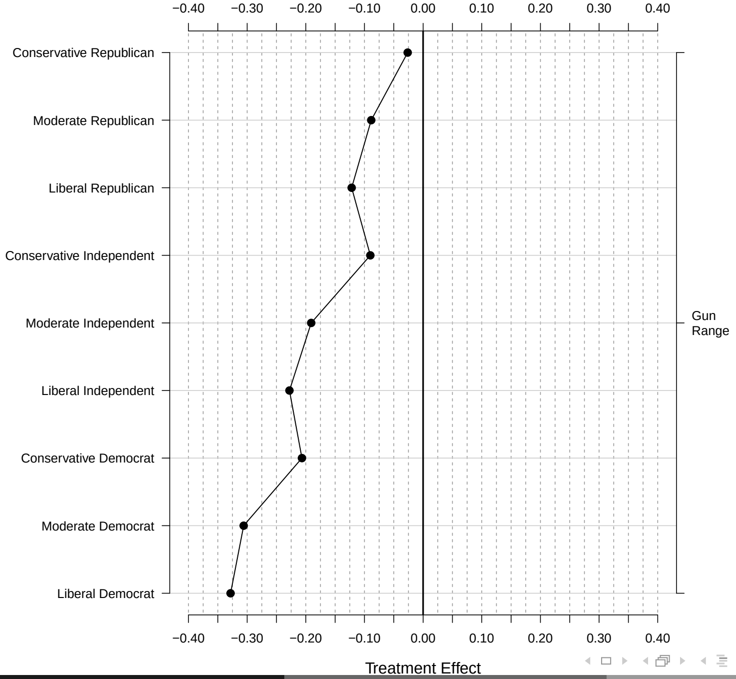

Conservative Republican
Moderate Republican
Liberal Republican
Conservative Independent
Moderate Independent
Liberal Independent
Conservative Democrat
Moderate Democrat
Liberal Democrat
Roads
Police
Fire Department
Gun Range
Parks
Planned Parenthood
−0.40 −0.35 −0.30 −0.25 −0.20 −0.15 −0.10 −0.05 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
Treatment Effect



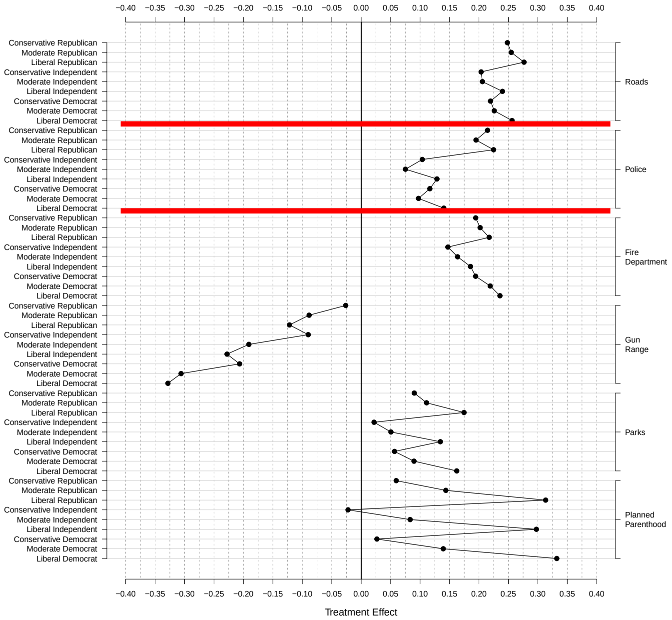
−0.40 −0.30 −0.20 −0.10 0.00 0.10 0.20 0.30 0.40
Conservative Republican
Moderate Republican
Liberal Republican
Conservative Independent
Moderate Independent
Liberal Independent
Conservative Democrat
Moderate Democrat
Liberal Democrat
Planned Parenthood
Treatment Effect

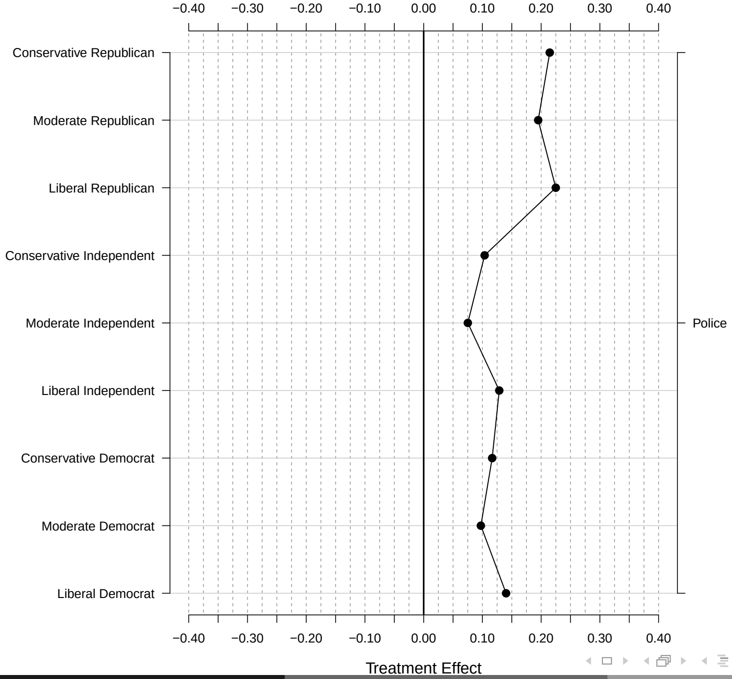


−0.40 −0.35 −0.30 −0.25 −0.20 −0.15 −0.10 −0.05 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
Conservative Republican
Moderate Republican
Liberal Republican
Conservative Independent
Moderate Independent
Liberal Independent
Conservative Democrat
Moderate Democrat
Liberal Democrat
Roads
Police
Fire Department
Gun Range
Parks
Planned Parenthood
Treatment Effect

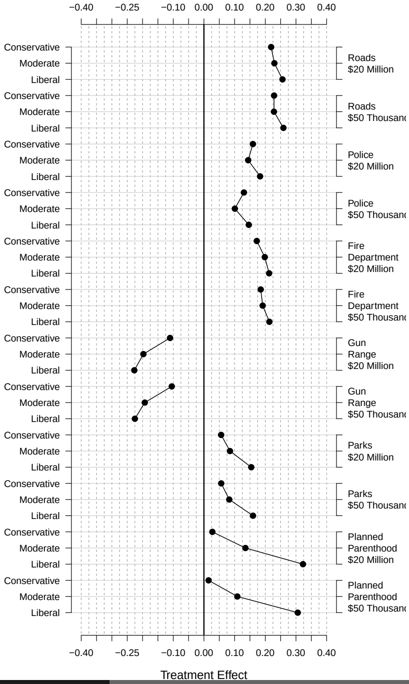


−0.40
−0.30
−0.20
−0.10
0.00
0.10
0.20
0.30
0.40
Conservative Republican
Moderate Republican
Liberal Republican
Conservative Independent
Moderate Independent
Liberal Independent
Conservative Democrat
Moderate Democrat
Liberal Democrat
Gun Range
Treatment Effect

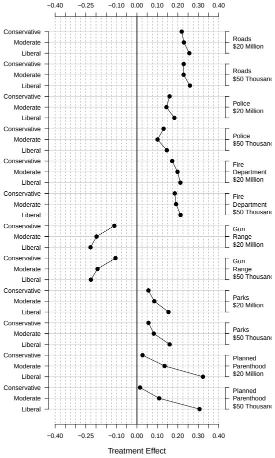

Treatment Effect
Roads
Police
Fire Department
Gun Range
Parks
Planned Parenthood
Conservative Republican
Moderate Republican
Liberal Republican
Conservative Independent
Moderate Independent
Liberal Independent
Conservative Democrat
Moderate Democrat
Liberal Democrat
−0.40 −0.35 −0.30 −0.25 −0.20 −0.15 −0.10 −0.05 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40



−0.40
−0.30
−0.20
−0.10
0.00
0.10
0.20
0.30
0.40
Conservative Republican
Moderate Republican
Liberal Republican
Conservative Independent
Moderate Independent
Liberal Independent
Conservative Democrat
Moderate Democrat
Liberal Democrat
Police
Treatment Effect



−0.40 −0.25 −0.10 0.00 0.10 0.20 0.30 0.40
Conservative
Moderate
Liberal
Roads
$20 Million
Roads
$50 Thousand
Police
$20 Million
Police
$50 Thousand
Fire
Department
$20 Million
Fire
Department
$50 Thousand
Gun
Range
$20 Million
Gun
Range
$50 Thousand
Parks
$20 Million
Parks
$50 Thousand
Planned
Parenthood
$20 Million
Planned
Parenthood
$50 Thousand
Treatment Effect



$\rightsquigarrow$ Constituents evaluate expenditures using qualitative information, rather than numerical facts

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects

R package $\rightsquigarrow$ implement methods, create synthetic observations, visualize results

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects

R package $\rightsquigarrow$ implement methods, create synthetic observations, visualize results

Ensembles as companion:

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects
R package $\rightsquigarrow$ implement methods, create synthetic observations, visualize results
Ensembles as companion:

- Better individual methods $\rightsquigarrow$ better ensembles

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects

`R` package $\rightsquigarrow$ implement methods, create synthetic observations, visualize results

Ensembles as companion:

- Better individual methods $\rightsquigarrow$ better ensembles
- Evaluate new methods $\rightsquigarrow$ more weight from ensemble

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects

R package $\rightsquigarrow$ implement methods, create synthetic observations, visualize results

Ensembles as companion:

- Better individual methods $\rightsquigarrow$ better ensembles
- Evaluate new methods $\rightsquigarrow$ more weight from ensemble
  1) Distinct

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects

R package $\rightsquigarrow$ implement methods, create synthetic observations, visualize results

Ensembles as companion:

- Better individual methods $\rightsquigarrow$ better ensembles
- Evaluate new methods $\rightsquigarrow$ more weight from ensemble
    1) Distinct
    2) Accurate

Ensembles to Estimate Heterogeneous Effects

Ensembles: prediction, classification, estimation of heterogeneous effects

R package $\rightsquigarrow$ implement methods, create synthetic observations, visualize results

Ensembles as companion:

- Better individual methods $\rightsquigarrow$ better ensembles
- Evaluate new methods $\rightsquigarrow$ more weight from ensemble
  1) Distinct
  2) Accurate

Ensembles $\rightsquigarrow$ leverage many contributions to build better estimates.