Machine Learning:
Introduction to Supervised Learning
Announcements

• Please fill out course evaluations!
• Can pick up exams after class
  • Extra exams will be in Gates 160
• Criteria posted online later today
  • All regrade requests **must** be made by this Friday
• Aubrey's Office Hours:
  • Wednesday: 12-5PM
  • Thursday: 12-5PM
  • Friday: 12-5PM (for regrade requests)
Buying Houses

- Given housing price data, how can we make generalizations from that data in order to predict housing prices?

<table>
<thead>
<tr>
<th>Square Feet</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>$100,000</td>
</tr>
<tr>
<td>800</td>
<td>$200,000</td>
</tr>
<tr>
<td>400</td>
<td>$120,000</td>
</tr>
<tr>
<td>550</td>
<td>$180,000</td>
</tr>
<tr>
<td>700</td>
<td>$250,000</td>
</tr>
<tr>
<td>900</td>
<td>$300,000</td>
</tr>
<tr>
<td>300</td>
<td>$120,000</td>
</tr>
<tr>
<td>500</td>
<td>$200,000</td>
</tr>
<tr>
<td>475</td>
<td>$120,000</td>
</tr>
</tbody>
</table>
Buying Houses

![Scatter plot showing the relationship between price and square feet of houses.](image)
Buying Houses

Price

Square Feet
Buying Houses

Graph showing the relationship between Price and Square Feet. The graph includes a line of best fit. The question mark at 600 indicates uncertainty regarding the value at that square footage.
Buying Houses

Price

Square Feet

600 ft²?
Buying Houses

Price vs. Square Feet

600 ft²?
Buying Houses

Price

Square Feet

$250,000

600 ft²?
Buying Houses

• The line we picked “looked good”, but how do we decide what line to pick?
Buying Houses

![Scatter plot showing the relationship between square feet and price. The larger the square footage, the higher the price tends to be.]
Buying Houses

Price vs. Square Feet
Buying Houses

Price vs. Square Feet

The graph shows the relationship between the price of houses and their square footage, with data points plotted and trend lines indicating a positive correlation.
Picking Lines

- One idea is to minimize the error between our data points and line we choose.
Buying Houses

Price

Square Feet
Buying Houses

Price

Square Feet
Picking Lines

• One idea is to minimize the error between our data points and line we choose.
  • Error = sum of the squared distances between the line and our data points
• How do formulate this mathematically?
Picking Lines

- The function for a line is:
  
  \[ f(x) = ax + b \]

- Let's represent our data points symbolically as:
  
  \[ \text{square footage} = x_n \]
  \[ \text{price} = y_n \]

- The error between one of our data points and our prediction for that data point is:
  
  \[ \text{error} = (y_n - f(x_n))^2 \]

  \[ \text{total error} = \sum (y_n - f(x_n))^2 \]
Picking Lines

• The overall error of our line is:

\[ total\ error = \sum (y_n - f(x_n))^2 \]

• Now that we have a function for our error, we can evaluate the quality of a line we choose

• This still doesn't tell us how to pick a line
Calculus to the Rescue!

- Answer: Some relatively straightforward calculus will give us a closed form solution for both \( b \) and \( a \).
  - We can now pick the “best” line!
    - Are we done? (Hint: NO)
Calculus to the Rescue!

- No! 3 questions remain:
Calculus to the Rescue!

• No! 3 questions remain:
  1) Does our data really fit a line?
Buying Houses

[Scatter plot with points representing price vs. square feet, showing a positive correlation.]

Price

Square Feet
Buying Houses

Price vs. Square Feet

- Data points represent houses.
- The red line indicates a trend in the data.
- The relationship suggests that as square footage increases, price also increases.
Calculus to the Rescue!

• No! 3 questions remain:

  1) Does our data *really* fit a line?
  2) Did we pick the right error function?
Buying Houses

![Graph showing the relationship between square feet and price.](image_url)
Buying Houses

![Graph showing the relationship between price and square feet. The graph has points scattered along a line, indicating a positive correlation.](image-url)
Buying Houses

The squared error of this element will have a huge impact on the line.
Visualizing Squared Error

Price

Square Feet
Visualizing Squared Error
Visualizing Squared Error
Visualizing Squared Error

Squared error puts a lot of emphasis on outliers.
Calculus to the Rescue!

No! 3 questions remain:

1) Does our data really fit a line?
2) Did we pick the right error function?
3) Is square footage the best way to predict housing prices?
Buying Houses

Price vs. Square Feet
Buying Houses

Price vs. Square Feet

- Scatter plot showing the relationship between price and square footage for houses.
Buying Houses

![Graph showing the relationship between price and square feet for houses. The graph has points scattered on a grid, with price on the y-axis and square feet on the x-axis. The points suggest a positive correlation.]
Supervised Learning

- **Supervised Learning** is a form of Machine Learning in which you have labeled training data and you train a model to predict these labels.

Alzheimer's?  
**YES/NO**
Other Forms of Learning

- **Unsupervised Learning**: Given training data *without labels* tell me something interesting about the data.
Other Forms of Learning

- **Reinforcement Learning**: Given an environment and an objective function, learn a model to make choices in the environment.

http://www.dacya.ucm.es/jam/images/MountainCar.jpg
When posed with a problem in Machine Learning you have to answer 3 questions:

1) How do I represent my data?
   - Square footage? Number of bathrooms? Quality of nearby schools?

2) What type of function do I want to learn?
   - Line? Parabola?
   - What is our Error function?

3) How do I optimize my model?
   - Gradient Ascent? Solve directly?
   - Matrix factorization?
Other Questions

- What I don't have very much training data?
  - What if you have extra “unlabeled” data?
    - e.g. MRI scans of undiagnosed patients
  - What if you have other types of supervision?
    - e.g. I know these two instances belong to the same cluster, I just don't know which cluster they belong to.
- What if you know something about the function you're going to learn?
  - e.g. Most of the features are not useful.
Introduction to Perceptrons
Handwriting Analysis

- Train a computer to recognize images of handwritten numbers 0 – 9.
- Large training and test set available (MNIST Handwritten Digit Database)
Classification Problem

• The problem we are considering is a classification problem: given an image, we need to figure what class it belongs to (0, 1, 2, 3, etc.).

• Just to be concrete, we're going to write a program with:
  • Input: Image of a handwritten digit
  • Output: Prediction for what digit the image is
Question 1: How do we represent an image?
<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_6$</td>
<td>$x_{11}$</td>
<td>$x_7$</td>
<td>$x_{12}$</td>
<td>$x_8$</td>
<td>$x_{13}$</td>
</tr>
<tr>
<td>$x_{11}$</td>
<td>$x_6$</td>
<td>$x_{16}$</td>
<td>$x_{12}$</td>
<td>$x_{17}$</td>
<td>$x_{13}$</td>
</tr>
<tr>
<td>$x_{16}$</td>
<td>$x_{11}$</td>
<td>$x_{16}$</td>
<td>$x_{18}$</td>
<td>$x_{17}$</td>
<td>$x_{19}$</td>
</tr>
<tr>
<td>$x_{21}$</td>
<td>$x_{22}$</td>
<td>$x_{23}$</td>
<td>$x_{24}$</td>
<td>$x_{25}$</td>
<td></td>
</tr>
</tbody>
</table>
Question 2: What function are we trying to optimize?
\[ f(x_1, x_2, \ldots, x_n) = w_0 + \sum x_i w_i \]
Question 3: How do we optimize our perceptrons?
How do we choose good values for \( w_0 \ldots w_n \)?
One Approach

• **Train** the perceptron on valid data.
One Approach

- **Train** the perceptron on valid data.
- For each data point (image of a digit):
One Approach

- **Train** the perceptron on valid data.
- For each data point (image of a digit):
  - Ask the perceptron what it thinks.
One Approach

- **Train** the perceptron on valid data.
- For each data point (image of a digit):
  - Ask the perceptron what it thinks.
  - If correct, do nothing.
One Approach

- **Train** the perceptron on valid data.
- For each data point (image of a digit):
  - Ask the perceptron what it thinks.
  - If correct, do nothing.
  - Otherwise, nudge $w_0 \ldots w_n$ in the right direction.
One Approach

- **Train** the perceptron on valid data.
- For each data point (image of a digit):
  - Ask the perceptron what it thinks.
  - If correct, do nothing.
  - Otherwise, nudge $w_0 \ldots w_n$ in the right direction.
- Repeat until number of errors is “small enough.”
One Approach

- **Train** the perceptron on valid data.
- For each data point (image of a digit):
  - Ask the perceptron what it thinks.
  - If correct, do nothing.
  - Otherwise, nudge $w_0 \ldots w_n$ in the right direction.
- Repeat until number of errors is “small enough.”
- Question: What kind of mistakes can we make?
False Positive
False Positive

\[ x_1 \]
\[ x_2 \]
\[ x_3 \]
\[ x_4 \]
\[ x_5 \]
\[ x_6 \]
\[ \ldots \]
\[ x_n \]
\[ +1 \]

\[ w_0 \]
\[ w_1 \]
\[ w_2 \]
\[ w_3 \]
\[ w_4 \]
\[ w_5 \]
\[ w_6 \]

\[ > 0 \]

YES
False Positive

\[ w_1 - \alpha x_1 \]
\[ w_2 - \alpha x_2 \]
\[ w_3 - \alpha x_3 \]
\[ w_4 - \alpha x_4 \]
\[ w_5 - \alpha x_5 \]
\[ w_6 - \alpha x_6 \]
\[ \vdots \]
\[ w_n - \alpha x_n \]
\[ w_0 - \alpha x_0 \]

If the sum is greater than 0, then it is a False Positive.
False Negative
False Negative
False Negative

\[ w_0 + \alpha x_0 \]
\[ w_1 + \alpha x_1 \]
\[ w_2 + \alpha x_2 \]
\[ w_3 + \alpha x_3 \]
\[ w_4 + \alpha x_4 \]
\[ w_5 + \alpha x_5 \]
\[ w_6 + \alpha x_6 \]
\[ \ldots \]
\[ w_n + \alpha x_n \]

\[ \leq 0 \]
Visualizing w
Visualizing w
Visualizing w
Visualizing w
Visualizing $w$
Visualizing $w$
Visualizing $w$
A Nice Math Trick

- For false positives, set $w_i = w_i - \alpha x_i$.
- For false negatives, set $w_i = w_i + \alpha x_i$.
- For correct answers, set $w_i = w_i$.
- Let “YES” be 1 and “NO” be 0.
- Consider the difference between actual answer and perceptron guess:
  - False positive: Actually NO, we say YES. Difference is -1.
  - False negative: Actually YES, we say NO. Difference is +1.
  - Correct answer: Both YES or both NO. Difference is 0.
- General update rule: $w_i = w_i + \alpha (\text{real} - \text{guess}) x_i$. 
Perceptron Learning Algorithm

- Start with a random guess of each $w_i$.
- Repeat until perceptron is sufficiently accurate:
  - Choose a training example $(x_0, x_1, ..., x_n)$.
  - Let $\text{real}$ be the real answer, $\text{guess}$ be the perceptron's guess.
  - For each $i$, set $w_i := w_i + \alpha (\text{real} - \text{guess}) x_i$
Visualization of $w$!
Combining Perceptrons

This is called a neural network.
Perceptrons

• How “good” are perceptrons?
  • Not very...they have lots of issues.
  • Better algorithms exist for classification

• Chose to cover them because they don't require much math.
Machine Learning

Exam Statistics

- Median: 35
- First Quartile: 24
- Third Quartile: 42