Prize4Life: Predicting Disease Progression in ALS

Lester Mackey

November 22, 2013

Joint work with Lilly Fang

Special thanks to Neta Zach and Robert Küffner

Goals of the Talk

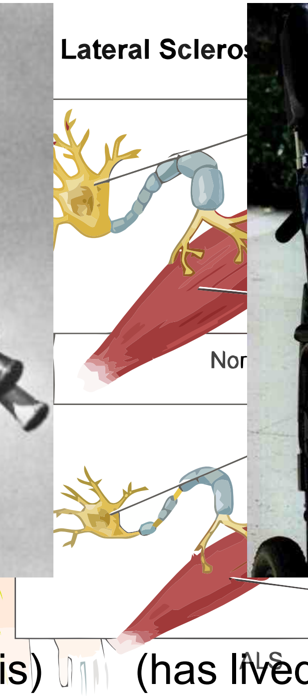
- Bring awareness to a fatal disease
  - Amyotrophic lateral sclerosis (ALS)
- Present an example of crowdsourced science
  - $50,000 ALS Prediction Prize4Life Challenge
- Introduce you to a rich data source
  - 8500 patient PRO-ACT database
- Highlight interesting (open) statistical questions

What is ALS?

- **Amyotrophic lateral sclerosis** or **Lou Gehrig's Disease**
  - A neurodegenerative disease that targets motor neurons
  - Leads to muscle atrophy, paralysis, and ultimately death
  - 100% fatal, typically within 3-5 years, but not always

Fast progressor

Lou Gehrig
(died within 2 years of diagnosis)



Slow progressor

Stephen Hawking
(has lived with the disease for 50 years)

Prize4Life

- 2004: 29-year-old Avi Kremer diagnosed with ALS
- 2006: Founded ALS non-profit PRIZE4LIFE
  - **Goal:** Accelerate development of treatment for ALS

Avi, 9 months after diagnosis

Avi, 2011, receiving Israeli PM award for Entrepreneurship and Innovation

Prize4Life: Incentives for Innovation

- $1M ALS Biomarker Prize, 2006-2011
  - **Goal:** Inexpensive, sensitive tool for monitoring disease progression and treatment efficacy
- $1M ALS Treatment Prize, 2008-Present
  - **Goal:** Therapy increasing lifespan of ALS mice by 25%
- $50K ALS Prediction Prize, 7/2012-10/2012
  - **Goal:** Predict rate of disease progression in ALS patients
    - Distinguish the slow progressors from the fast

Questions

- **What** do we mean by disease progression?
- **Why** is progression prediction valuable?
- **How** can we hope to predict progression accurately?

Predicting ALS Progression: What?

- **ALS Functional Rating Scale (ALSFRS)**
  - Measure of patient functionality, ranging from 0-40
  - Based on 10 questions regarding everyday activity:
    - Speaking, respiration, climbing stairs, dressing, writing, …
    - Activity score of 4 is normal, 0 is complete inability
  - Slow progressor loses 0-3 points per year
  - Fast progressor can lose 20

	Speech	Respira.	Saliv.	Swall.	Handwr	Cutting	Dress.	Turn.	Climb.	Walk.	Total
Visit 0	3	4	3	3	4	4	3	4	4	4	**36**
Month 1	3	4	3	3	4	4	3	4	4	4	**36**
Month 2	3	4	2	3	4	4	3	4	4	4	**35**
Month 3	3	4	2	3	4	4	3	4	4	3	**34**

State of Progression Prediction

Clinical Presentation:

- A 69 year old Caucasian female 19 months after diagnosis
- Bulbar onset (degeneration in muscles controlling speaking/swallowing)
- Weight stable and normal

	Speech	Respira.	Saliv.	Swall.	Handwr	Cutting	Dress.	Turn.	Climb.	Walk.	Total
Visit 0	3	4	3	3	4	4	3	4	4	4	**36**
Month 1	3	4	3	3	4	4	3	4	4	4	**36**
Month 2	3	4	2	3	4	4	3	4	4	4	**35**
Month 3	3	4	2	3	4	4	3	4	4	3	**34**

ALSFRS Scores

State of Progression Prediction

Clinical Presentation: Vitals and Lab Tests

	Respiratory rate	Pulse	Blood pressure
Visit 0	12	82	**150/80**
Month 1	18	81	**144/80**
Month 2	Missing	Missing	**Missing**
Month 3	18	92	**142/84**

	Urine pH	**Glucose**	Hemogl.	Bilirubin	**Trigly**	**Cholest**	K	Cl	Ca	Na	Phos	CO2	Albumin	Creatinine	(BUN)
Visit 0	7	**6.4**	133	9	1.25	**6.53**	4.1	104	2.35	139	1.36	26	46	62	7.85
Month 1	6	5.4	132	7	**2.35**	**6.11**	4.3	105	2.45	139	1.45	28	46	71	8.96
Month 2	7	**6.1**	127	7	**1.66**	**7.07**	4.6	106	2.38	140	1.23	26	47	71	8.43
Month 3	6	5.6	131	7	1.29	**6.53**	4.5	105	2.38	140	1.39	29	47	62	7.78

	Basophils	Eosinophils	Monocytes	Lymphocytes	Neutrophils
Visit 0	0.02	0.13	0.51	1.61	4.32
Month 1	0.03	0.19	0.52	1.61	4.05
Month 2	0.02	0.22	0.67	2.49	4.70
Month 3	0.07	0.21	0.71	2.35	4.37

State of Progression Prediction

Six expert ALS clinicians estimated change in ALSFRS over 9 months

Clinician	A	B	C	D	E	F	Average
Score	-3	-3	-4	-5	-6	-11	-5.33

Reality: The patient lost 12 points

Predicting ALS Progression: Why?

Why predict rate of disease progression?

- Helping clinicians
  - More accurate prognosis
  - Identifying predictive patient characteristics
    - Which lab tests worthwhile?
- Stratifying clinical trial patients
  - Less variability ⇒ fewer patients needed ⇒ less expensive, more interpretable clinical trials
  - Recent 1000 patient trial cost over $100 million
  - Using our algorithm, Prize4Life estimates a 20% reduction in patients needed to observe drug effect

Predicting ALS Progression: How?

The PRO-ACT Database

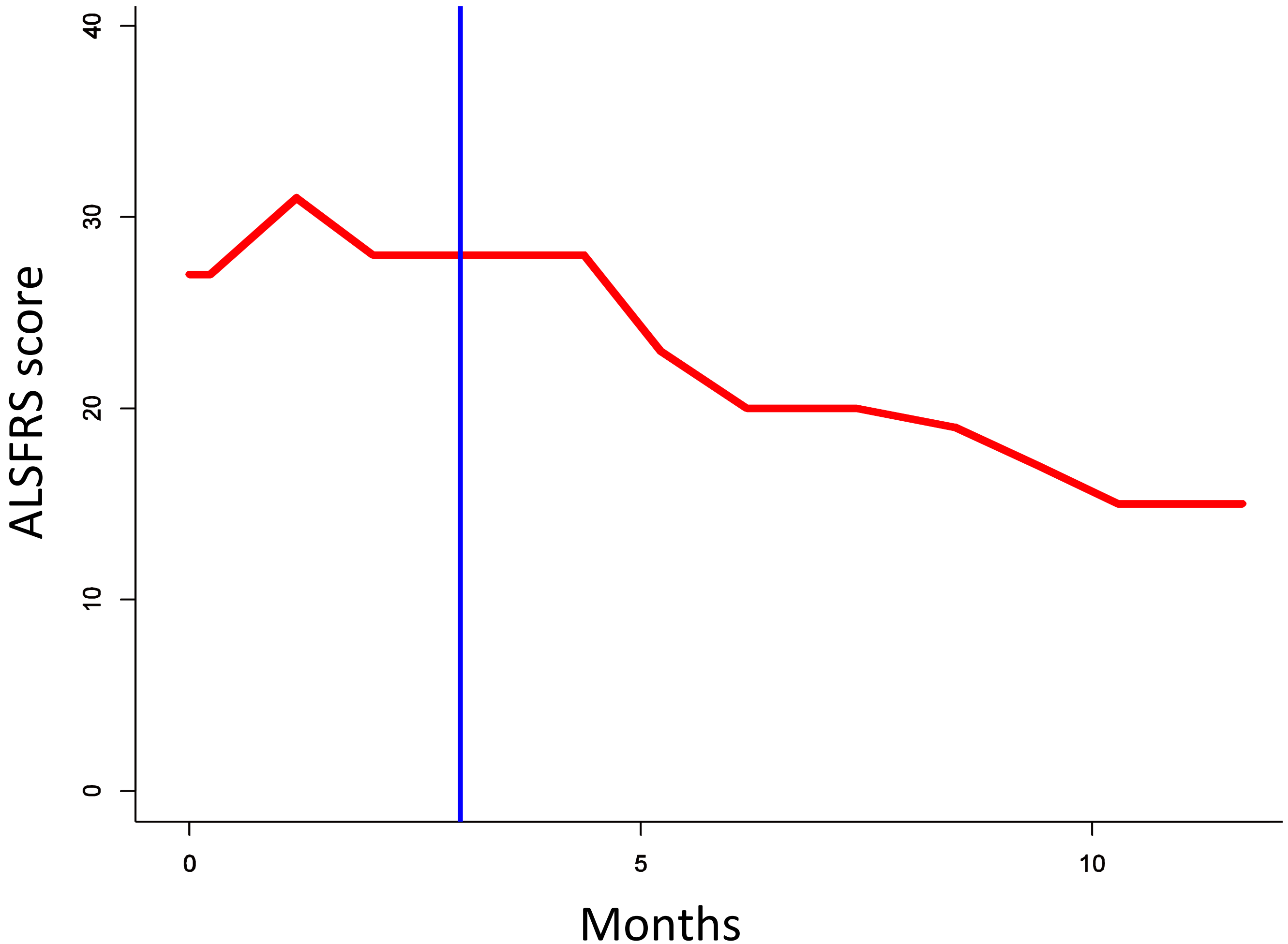
- **P**ooled **R**esource **O**pen-Access **A**LS **C**linical **T**rials
- 8500 de-identified patient records from completed clinical trials
  - Largest ALS patient data set ever assembled
  - Demographics, Medical and family history data
  - Functional measures (ALSFRS, lung capacity)
  - Vital signs (weight, height, respiratory rate)
  - Lab data (blood chemistry, hematology, and urinalysis)
- **Released to the public in Dec. 2012**

The ALS Prediction Prize

ALS Prediction Prize: Setup

- **The Contest Data**
  - 918 training patients
    - 12 months of data (demographic, ALSFRS, vital statistics, lab tests)
    - Time series: roughly monthly measurements, unequally spaced
  - 279 test patients
    - First 3 months of data available **at test time**
- **Challenge:** Given first 3 months of patient data, predict progression of ALS over subsequent 9 months
- **Measure:** ALS Functional Rating Scale (ALSFRS) score
  - Rate of progression = slope of ALSFRS score

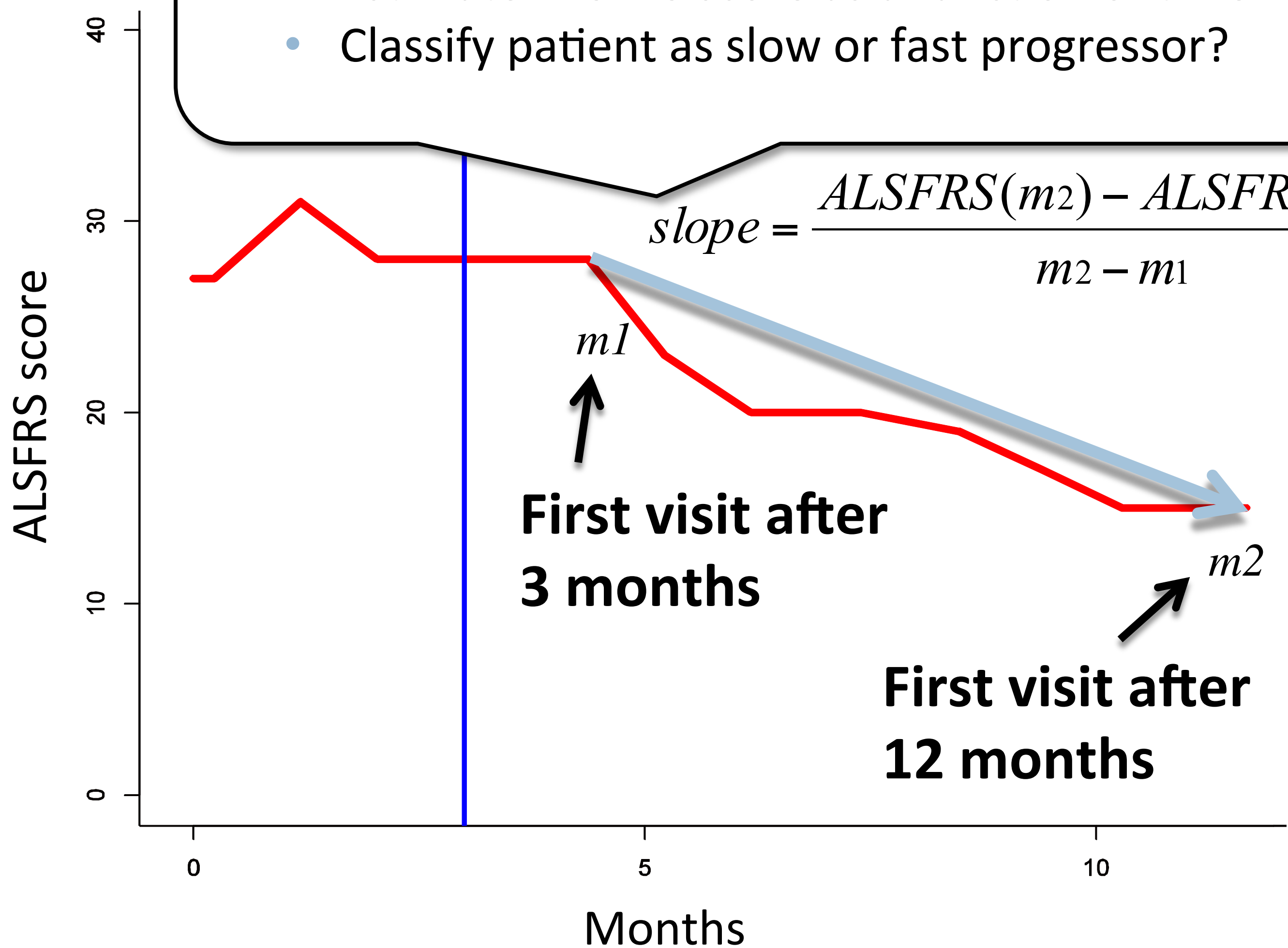
Target for Prediction



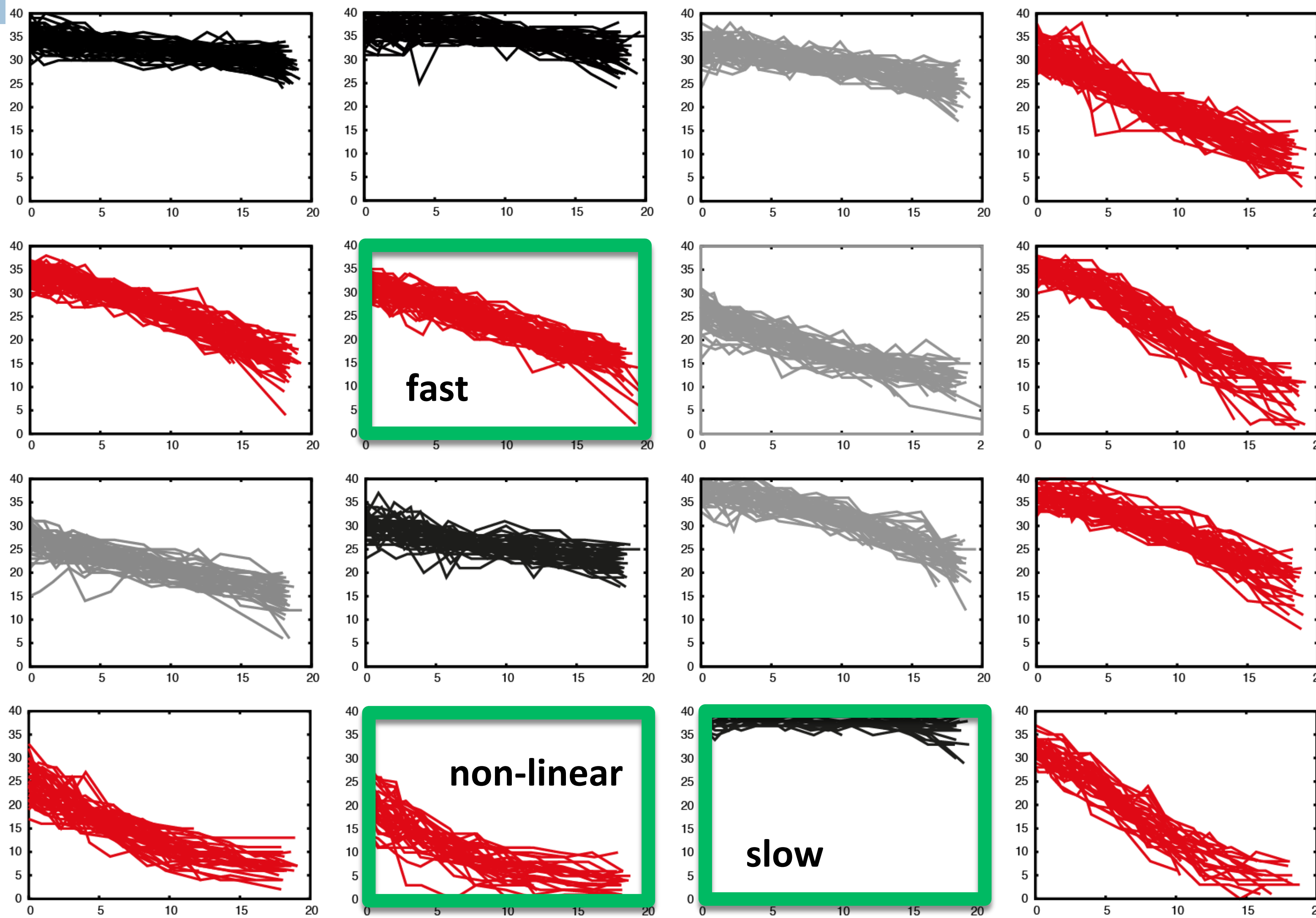
Target for Prediction

- **Issues:** Timing of future visits unknown; Slope unstable
- **Open Question: Better targets for prediction?**
    - Estimate ALSFRS score as a function of time?
    - Classify patient as slow or fast progressor?

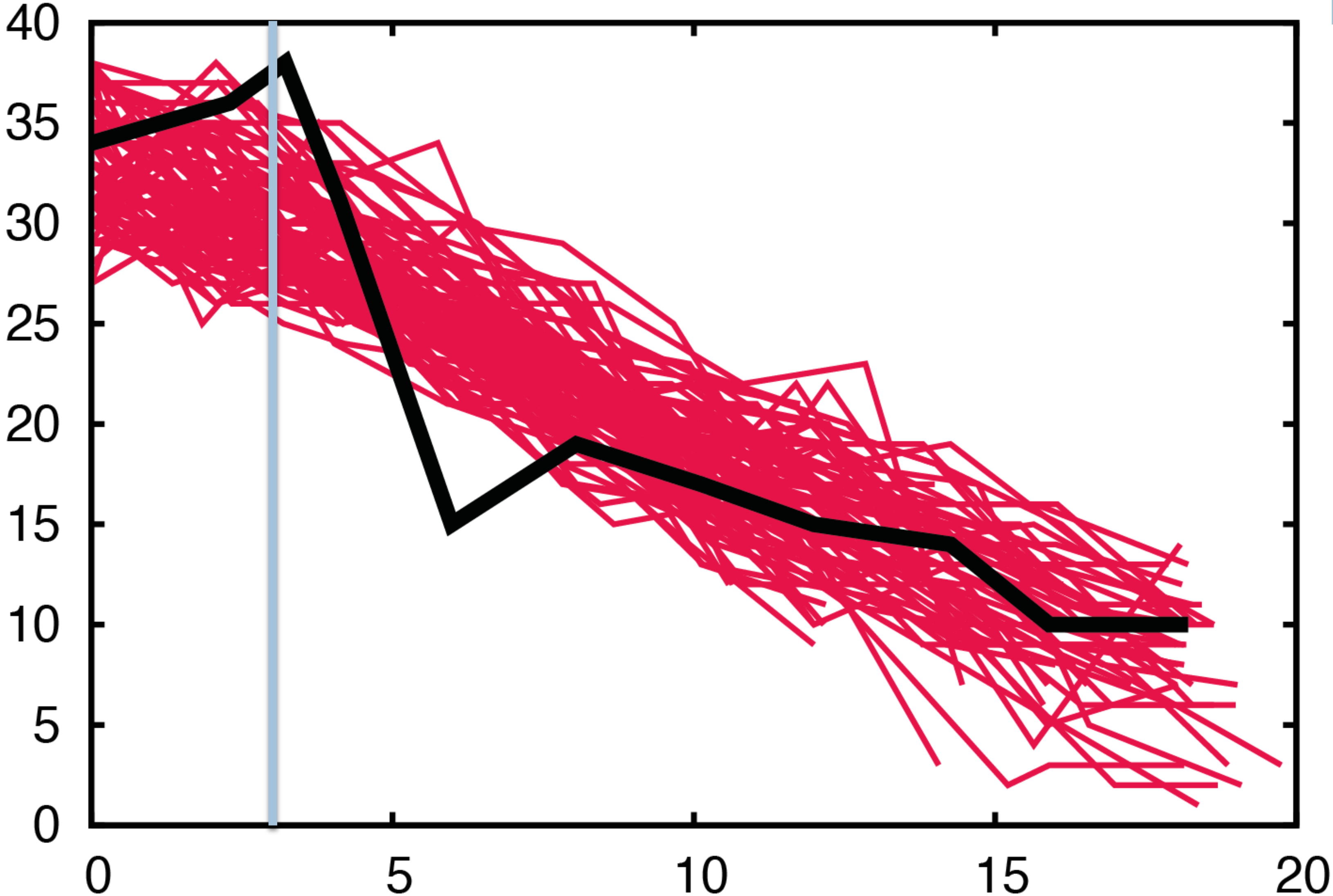
$$slope = \frac{ALSFRS(m_2) - ALSFRS(m_1)}{m_2 - m_1}$$



ALS Progression Types

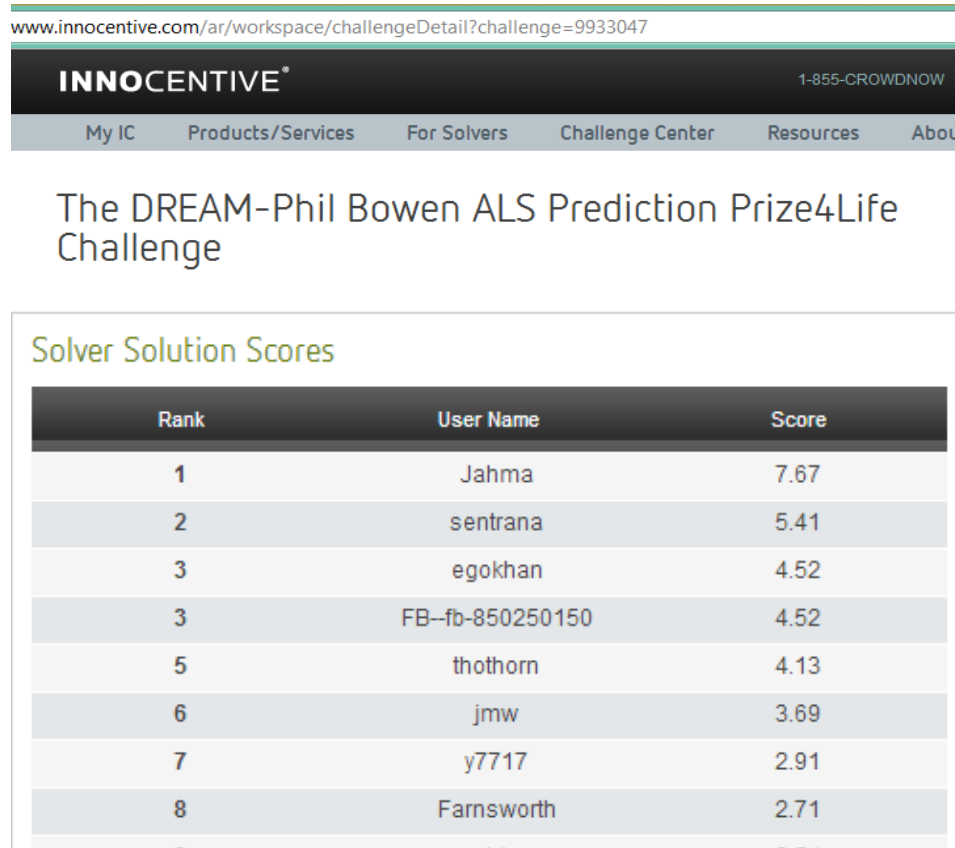


The Difficulty of Prediction



ALS Prediction Prize: Evaluation

- Contest run on **Innocentive** prize platform
  - Hosts science competitions
  - See also Kaggle, Challenge.gov
- Contestants **uploaded code** to Innocentive server
  - Code had to be written in R!
  - Max running time: 6 hours
- **Leaderboard** displayed error on test set
  - Max # submissions: 100
- **Error metric:** Root mean squared deviation (RMSD)


www.innocentive.com/ar/workspace/challengeDetail?challenge=9933047

INNOCENTIVE®

1-855-CROWDNOW

My IC | Products/Services | For Solvers | Challenge Center | Resources | Abou

The DREAM-Phil Bowen ALS Prediction Prize4Life Challenge

Solver Solution Scores

Rank	User Name	Score
1	Jahma	7.67
2	sentrana	5.41
3	egokhan	4.52
3	FB--fb-850250150	4.52
5	thothorn	4.13
6	jmw	3.69
7	y7717	2.91
8	Farnsworth	2.71

ALS Prediction Prize: Evaluation

- **Oct. 1, 2012:** Test set released to contestants
- **The Final Contest Data**
  - 918 training patients + 279 test patients
    - 12 months of data (demographic, ALSFRS, vital statistics, lab tests)
  - 625 validation patients determined prize winners
    - Data **never seen** by contestants, **no prior feedback** given
    - Tests ability to generalize to new patients

Our Approach

Featurization

- Static Data
- Time Series Data

Modeling and Inference

- Bayesian Additive Regression Trees

Post-hoc Evaluation

- BART Performance
- Feature Selection
- Model Comparison

Featurization

- **Goal:** Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features

Issue: Features manually specified by non-expert (me)

Open Question: Automatic featurization of longitudinal data?

Featurization

- **Goal:** Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features

- **Static Data**

49

- Demographics Age, Race, Sex
- ALS History Time from onset, Site of onset
- Family History Mother, Father, Grandmother, Uncle…
- ……………………
- ……………………

Categorical variables encoded as binary indicators

Featurization

- **Goal:** Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features

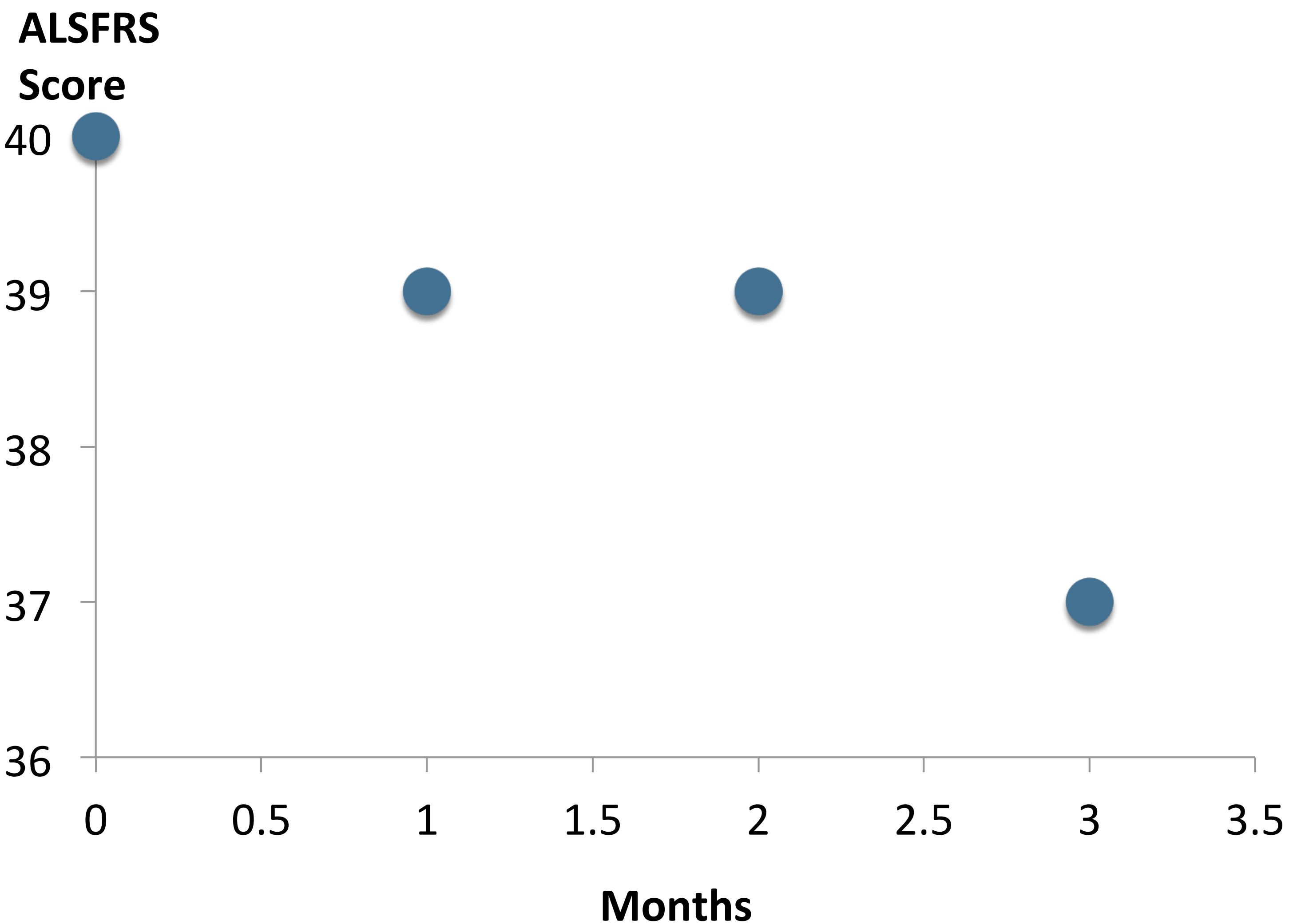
- **Time Series Data**
  - Repeated measurements of variables over time
    - ALSFRS question scores
    - Alternative ALS measures (forced and slow vital capacity)
    - Vital signs (weight, height, blood pressure, respiratory rate)
    - Lab tests (blood chemistry, hematology, urinalysis)
  - Number and frequency of measurements vary across patients

Featurization

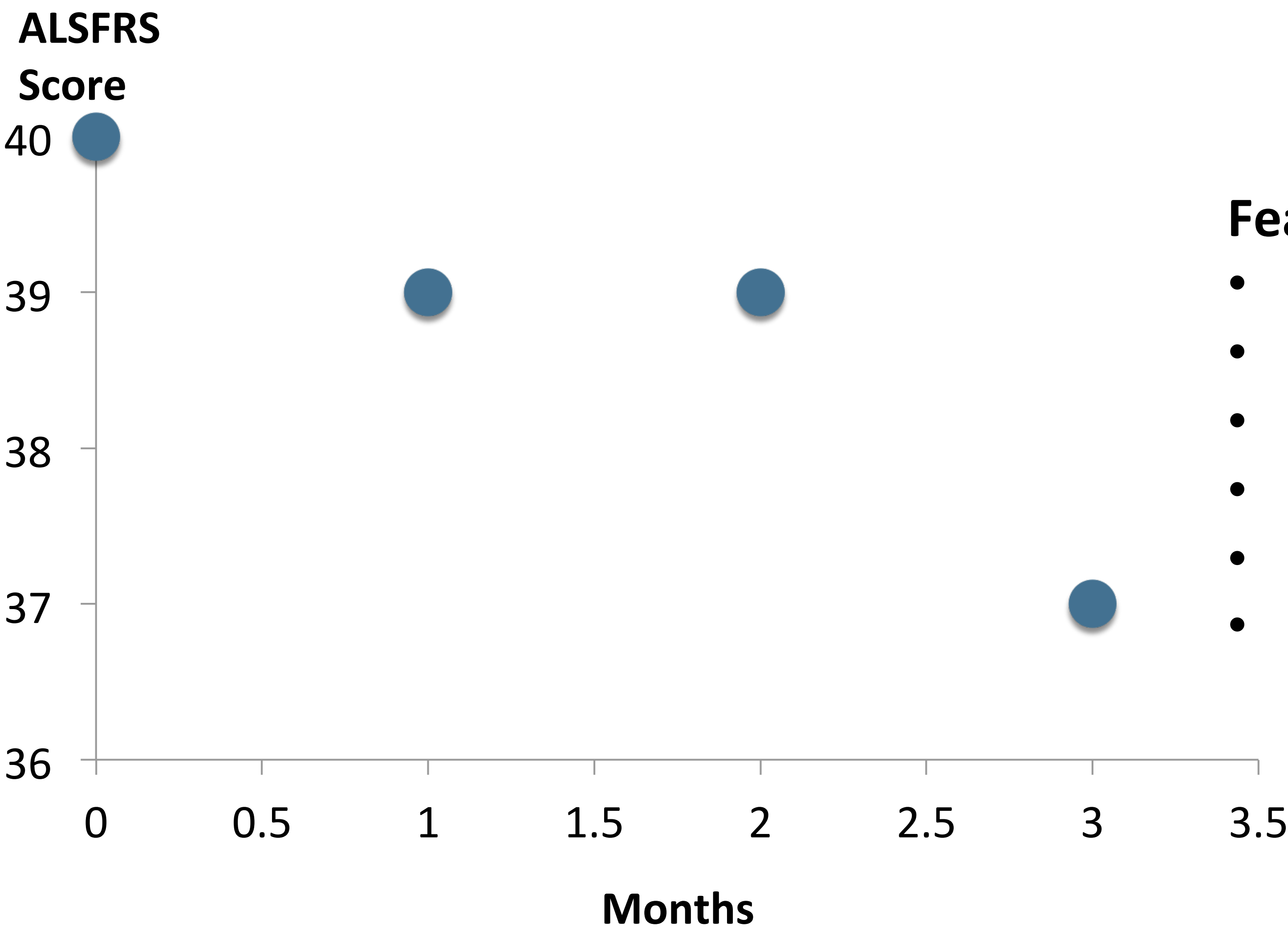
- **Goal:** Compact numeric representation of each patient
  - Features will serve as covariates in a regression model
  - Most extracted features will be irrelevant
  - Rely on model selection / methods robust to irrelevant features

- **Time Series Data**
  - Compute summary statistics from each time series
    - Mean value, standard deviation, slope, last recorded value, maximum value…
  - Compute pairwise slopes (difference quotients between adjacent measurements)
    - Induces a derivative time series
    - Extract same summary statistics

Featurizing Time Series Data



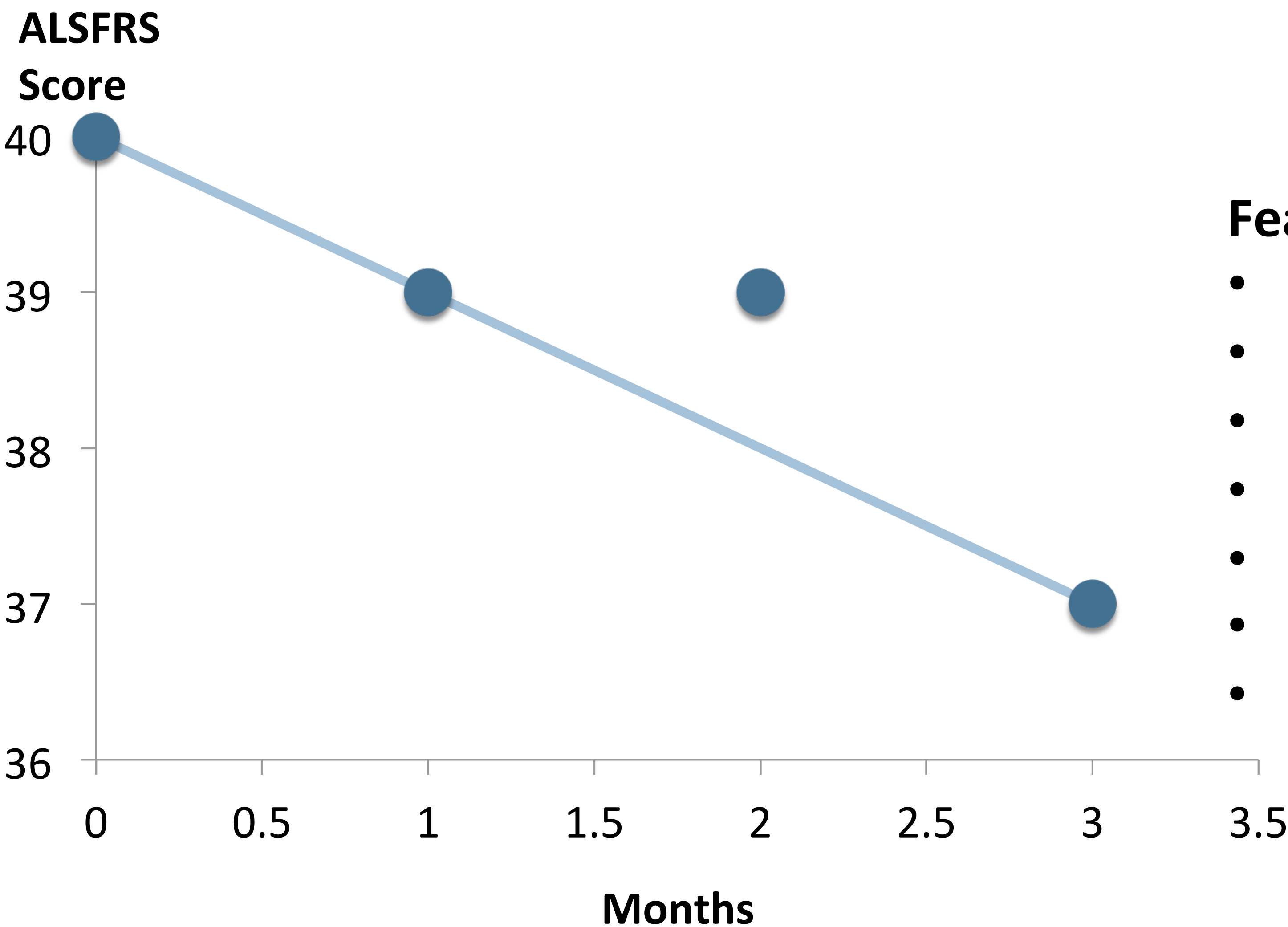
Featurizing Time Series Data



Features extracted

- Mean = 38.75
- SD = 0.816
- Max = 40
- Min = 37
- Last = 37
- etc.

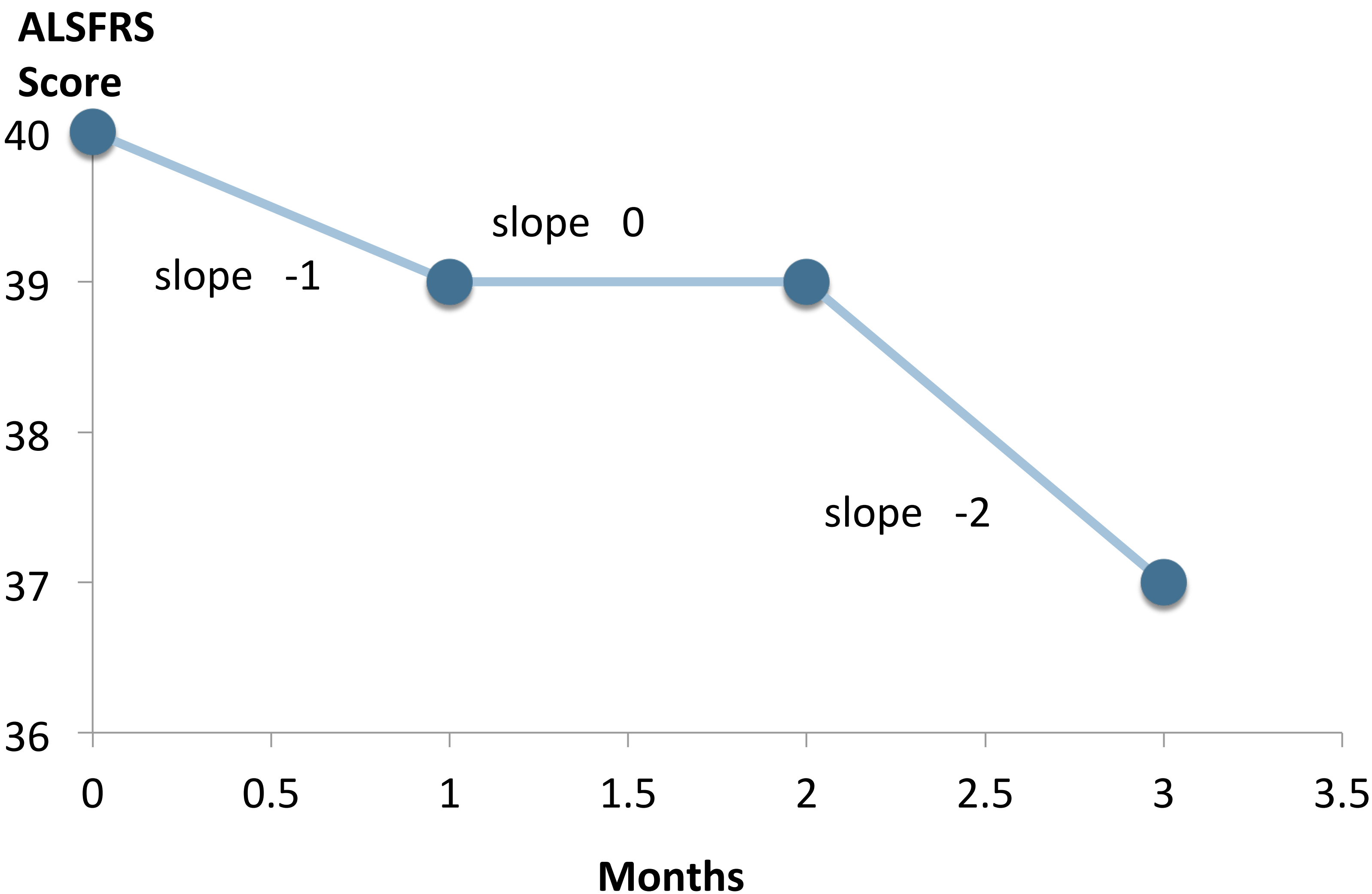
Featurizing Time Series Data



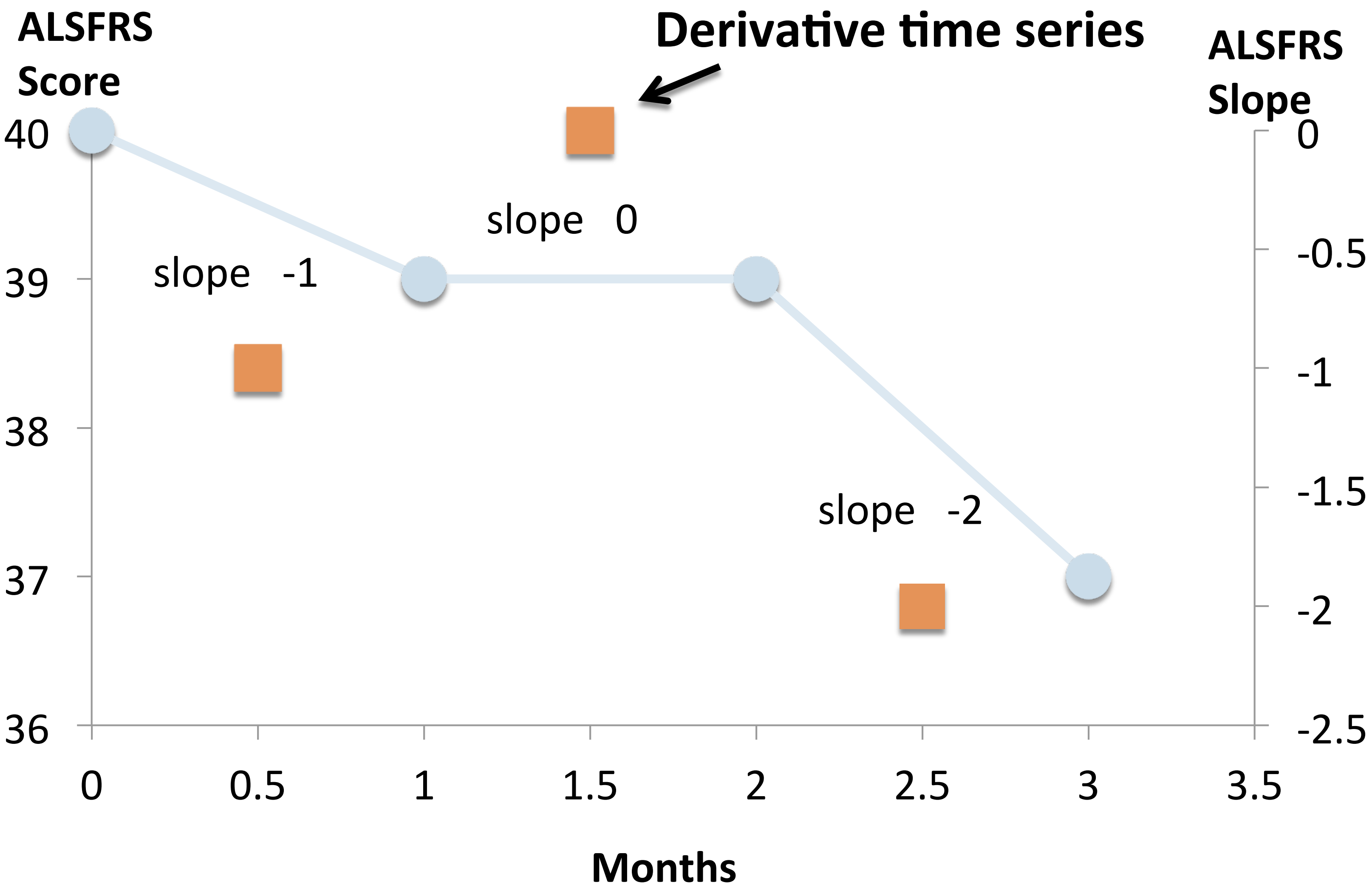
Features extracted

- Mean = 38.75
- SD = 0.816
- Max = 40
- Min = 37
- Last = 37
- Slope = -1
- etc.

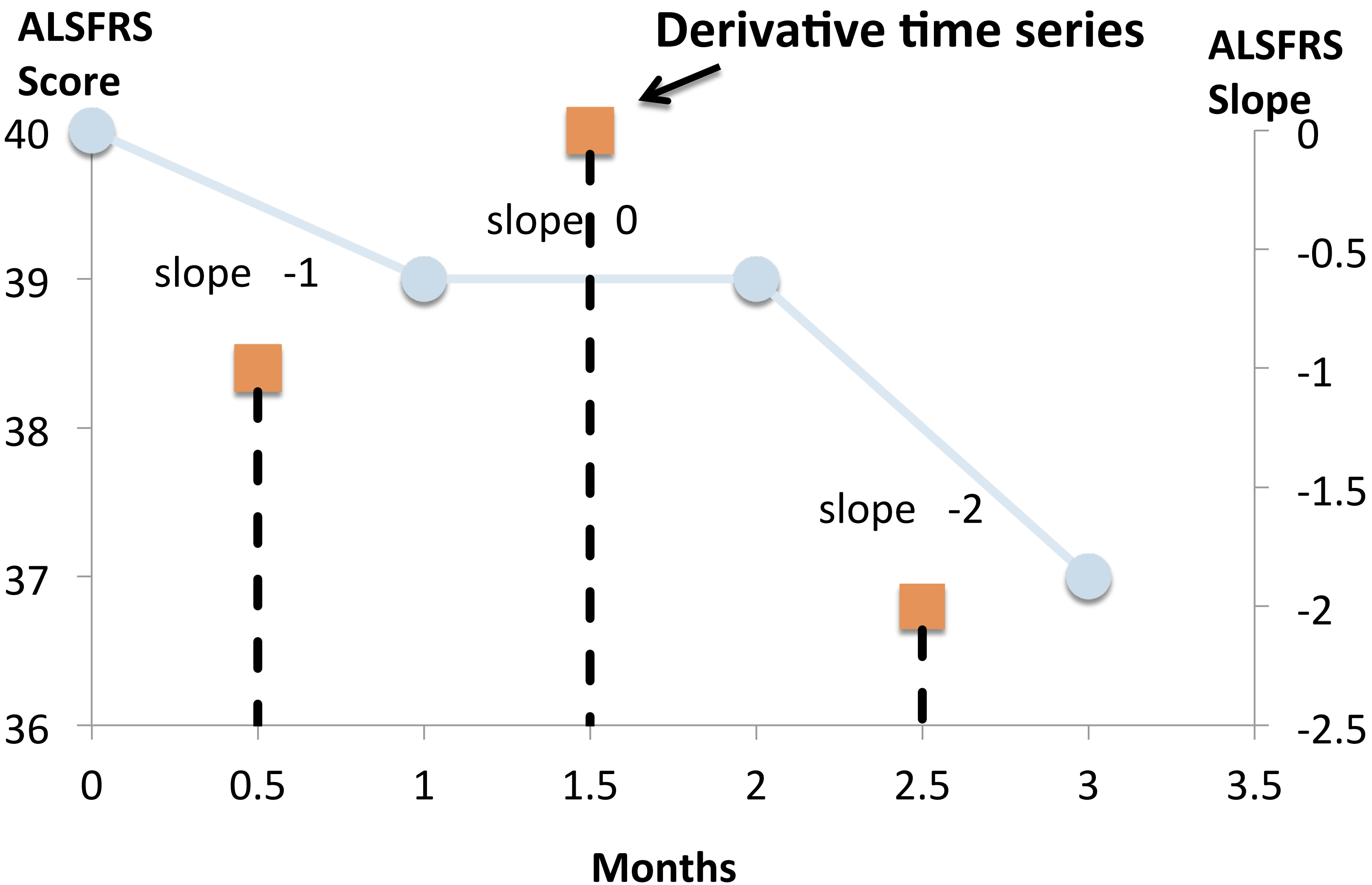
Featurizing Time Series Data



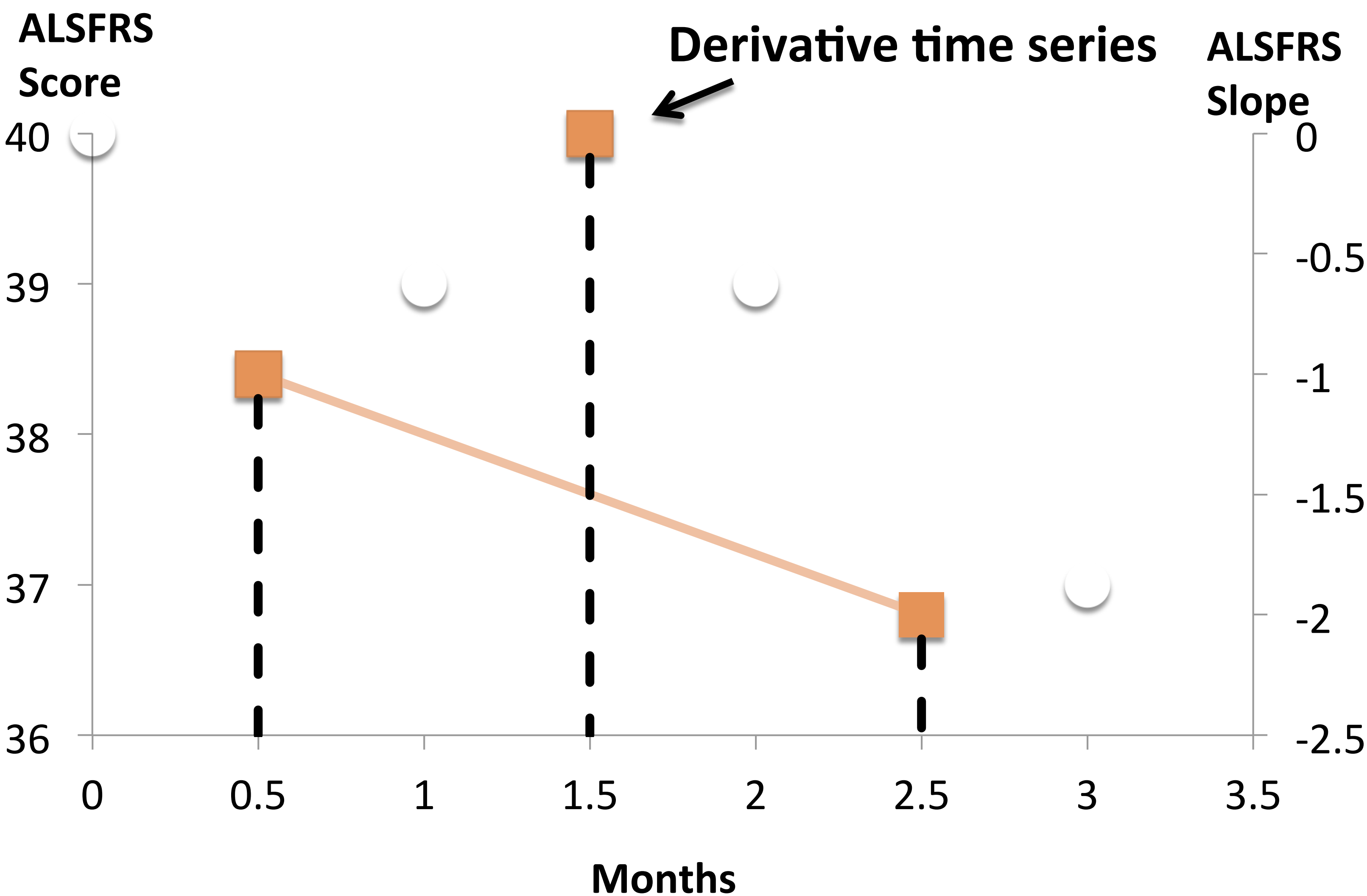
Featurizing Time Series Data



Featurizing Time Series Data



Featurizing Time Series Data



Features extracted

Mean = -1

SD = 1

Max = 0

Min = -2

Last = -2

Slope = -0.5

etc.

Featurizing Time Series Data

- **435** temporal features extracted
- **Problem:** Missing data
  - Average patient missing 10% of features
  - One patient missing 55% of features!
  - Missing values imputed using median heuristic

Room for improvement

- **Problem:** Outliers
  - Nonsense values: Number of liters recorded as `MDMD`
  - Units incorrectly recorded ⇒ Wrong conversions
  - Extreme values
    - Treated as missing if > 4 standard deviations from mean

Open Question: **Regression robust to (sparse) covariate outliers?**

Modeling and Inference

- **Regression model**

  Future ALSFRS Slope = f(features) + noise

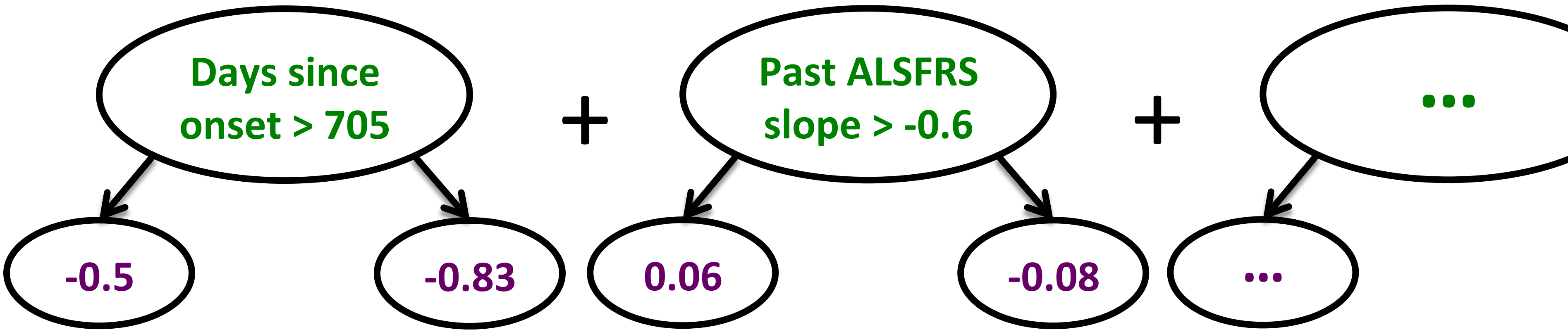
  Unknown regression function

- **Goal:** infer **f** from data
  - Bayesian: Place a prior on **f**, infer its posterior
  - Bonus: Uncertainty estimates for each prediction

- **What prior?**
  - Flexible and nonparametric
    - Avoid restrictive assumptions about functional form
  - Favor simple, sparse models
    - Avoid overfitting to irrelevant features

Bayesian Additive Regression Trees*

- f(features) = sum of **"simple" decision trees**



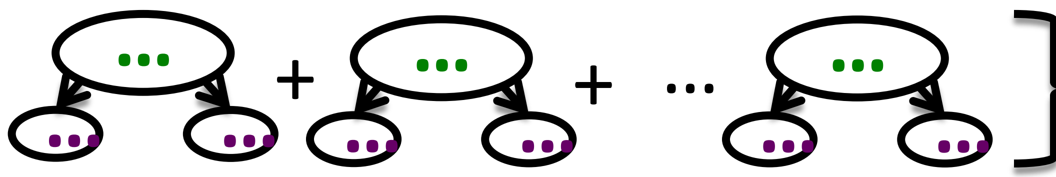
- **Simplicity** = tree depends on few features
  - Irrelevant features seldom selected
- Similar to frequentist ensemble methods
  - Boosted decision trees, random forests

*Chipman, George, and McCulloch (2010)

BART Inference

- **Estimating** **f**: Markov Chain Monte Carlo
  - R package ‘bart’ available on CRAN
  - 10,000 posterior samples: $\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4$, …

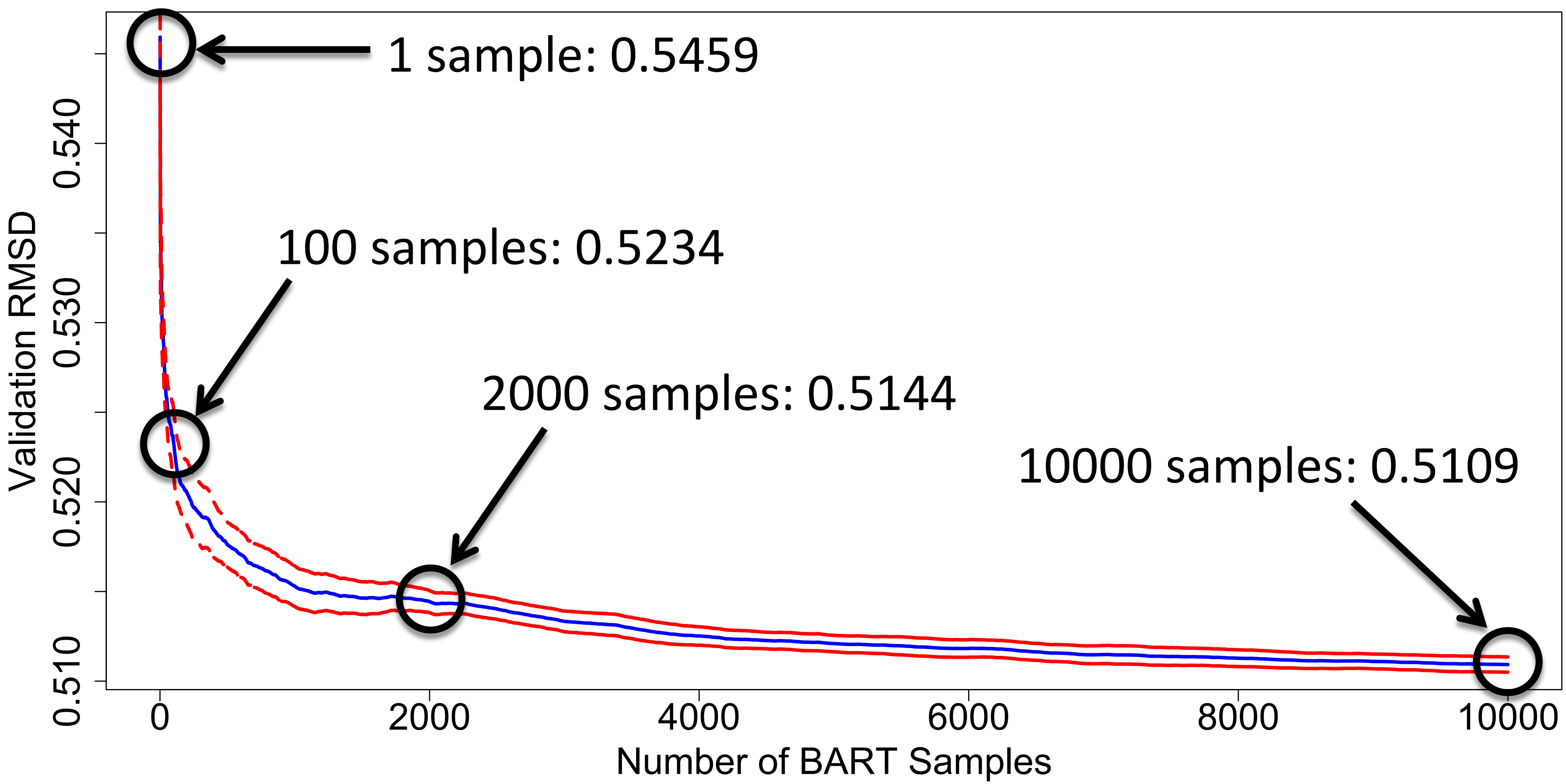
$\hat{f}_i =$



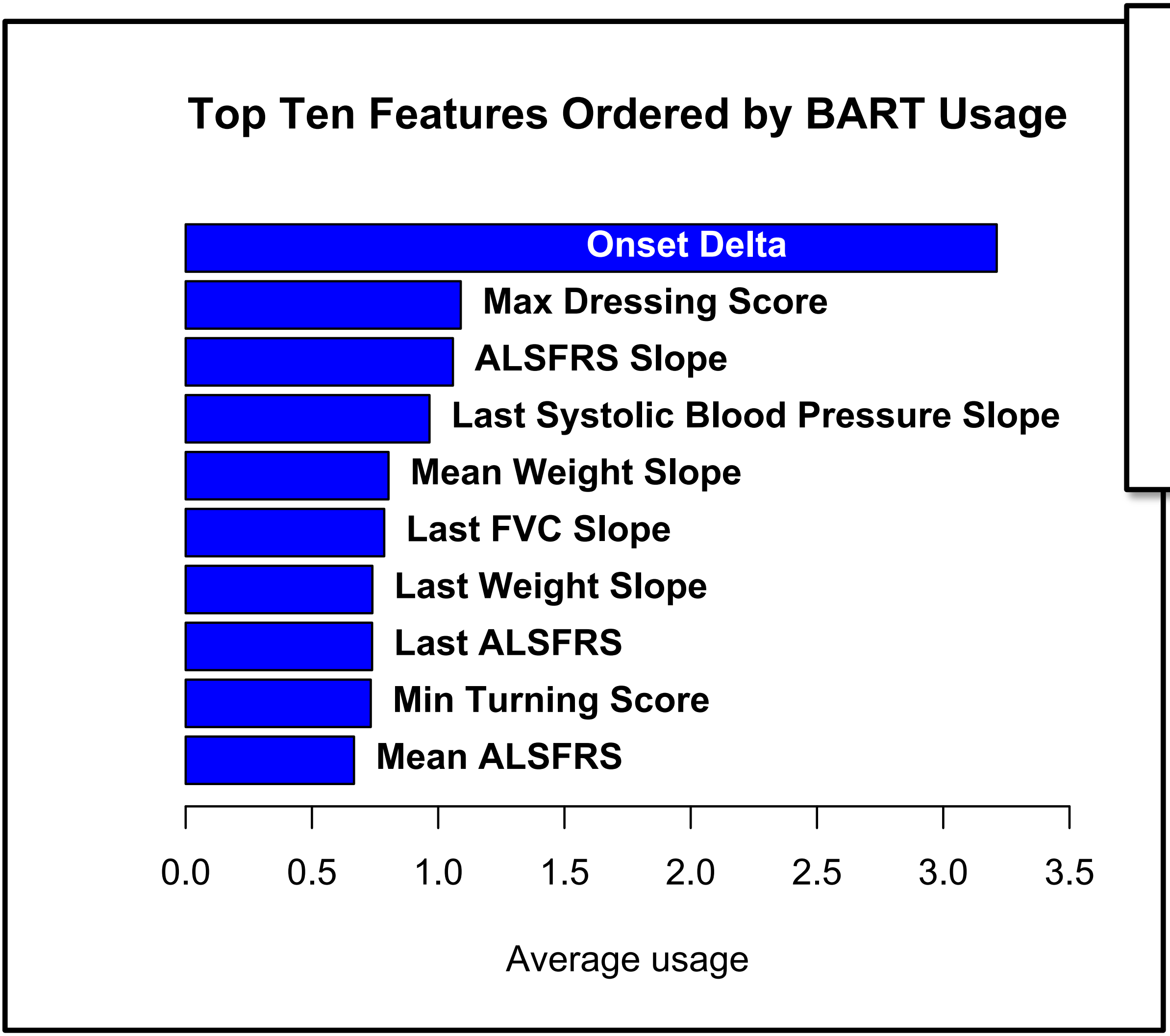
100 trees

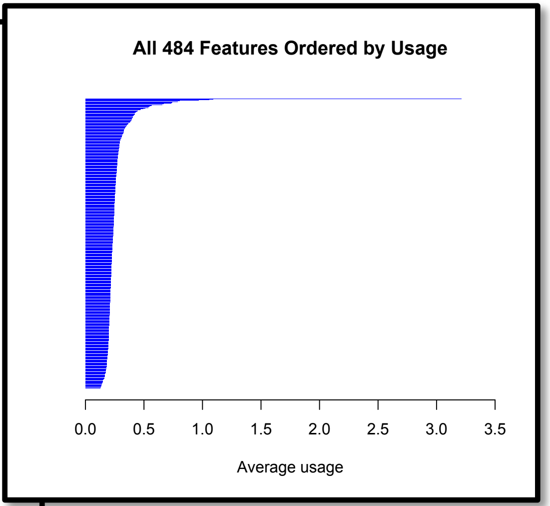
  - 10 minutes on MacBook Pro (2.5 GHz CPU, 4GB RAM)
- **Prediction:** Posterior mean
  - Average of $\hat{f}_1$(**features**), $\hat{f}_2$(**features**), $\hat{f}_3$(**features**), …
- **Variance reduction**
  - Average predictions of 10 BART models

Accuracy of BART Inference



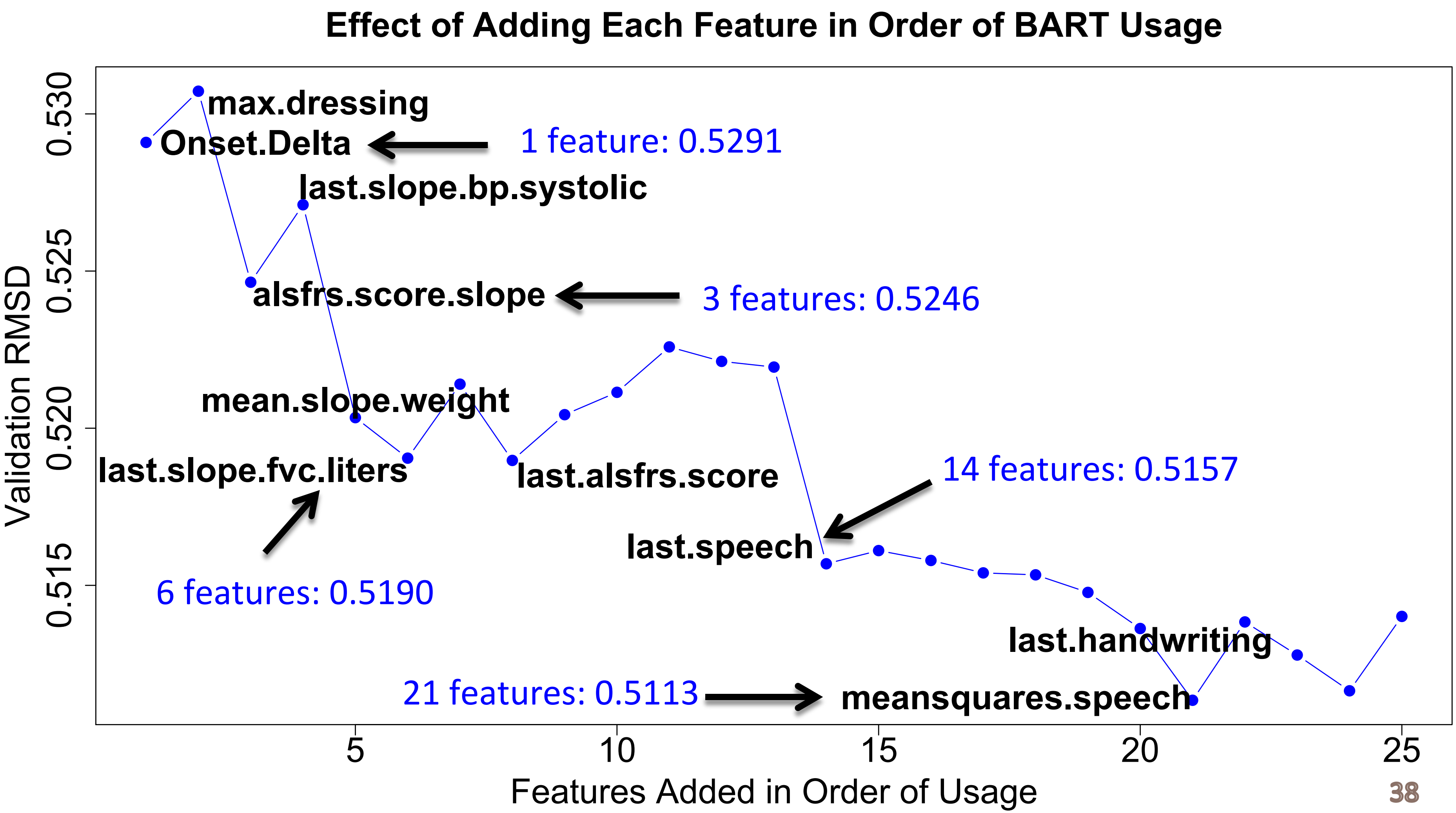
BART Feature Selection





- Many pairwise slope features
- Lab data excluded

BART on Feature Subsets



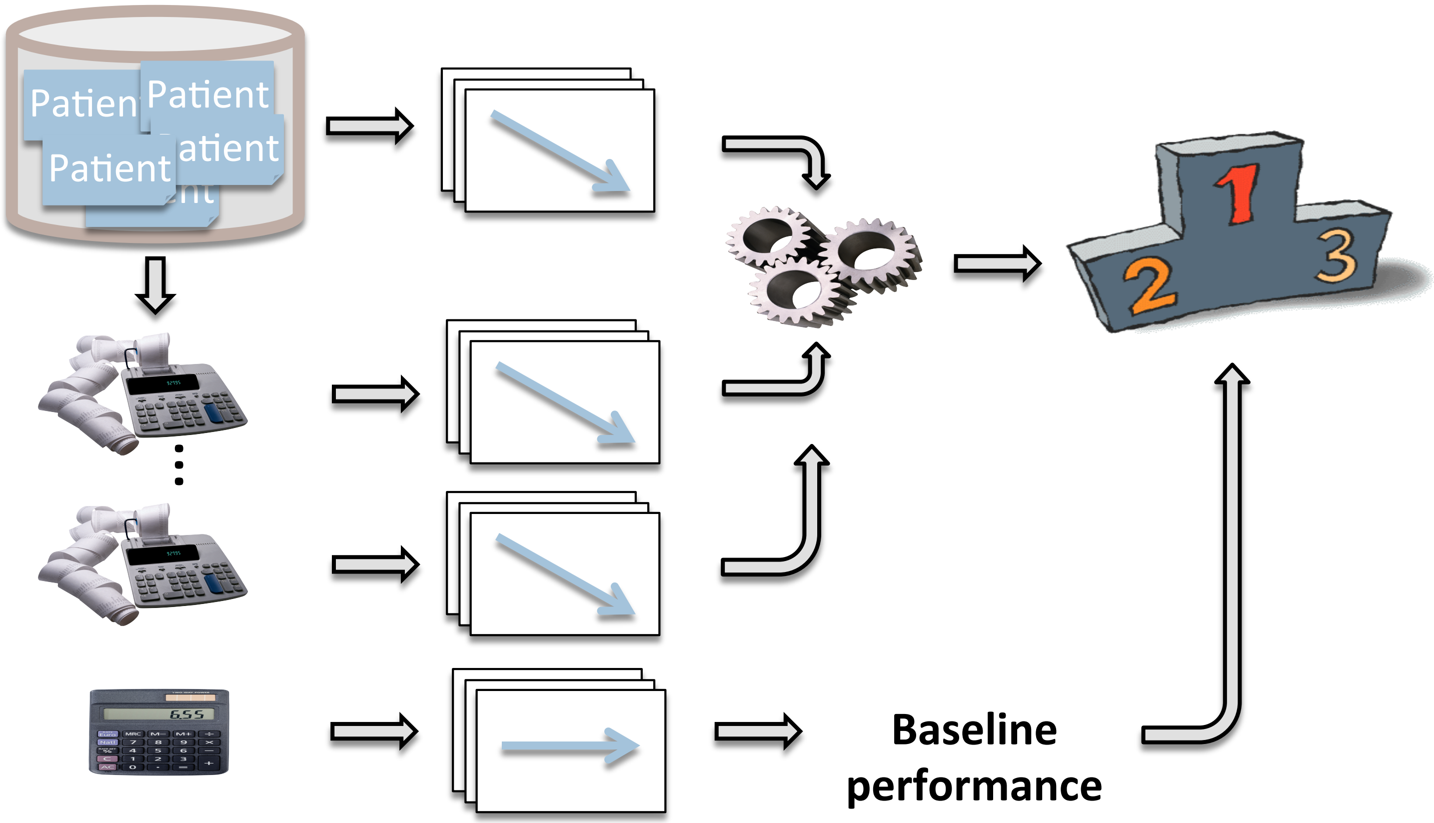
Model Comparison

How do other models perform using our feature set?

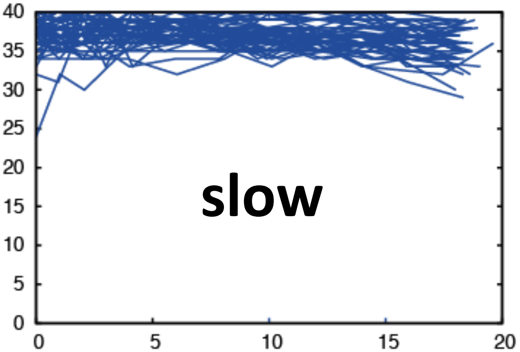
Model	Our RMSD (Test)	Our RMSD (Validation)	Competitor RMSD
Lasso Regression	0.5006	0.5287	-
Random Forests	0.5052	0.5120	0.52-0.53
Boosted Trees	0.4940	0.5118	-
BART	0.4860	0.5109	-

- **Additive decision tree** models especially effective
- **Featurization** was a main differentiator of competitors

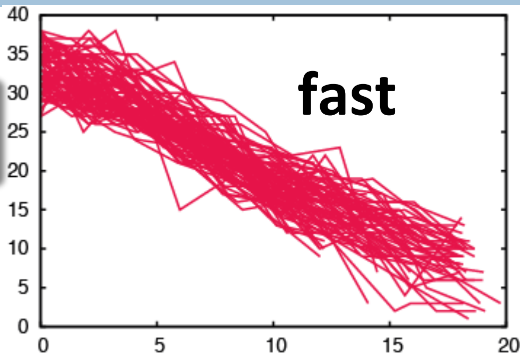
Contest Evaluation



RMSD: Slow vs. Fast Progressors



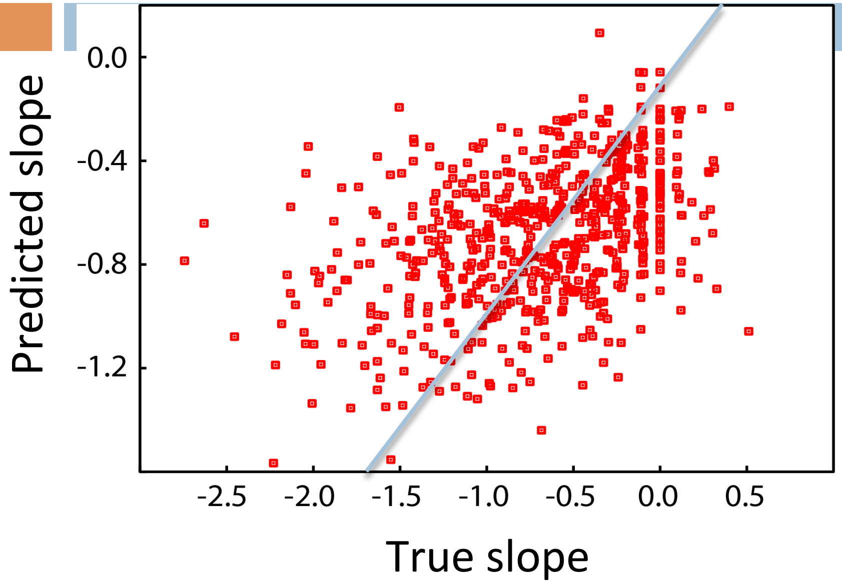
	all	slow	med	fast
1	0.51	0.43	0.29	0.78
2	0.52	0.43	0.30	0.79
3	0.52	0.40	0.30	0.84
4	0.53	0.42	0.31	0.83
5	0.53	0.44	0.30	0.82
6	0.53	0.38	0.34	0.88
7	0.57	0.46	0.26	0.91
8	0.57	0.47	0.36	0.88
9	0.89	0.92	0.61	1.04
10	1.30	1.04	1.43	1.67



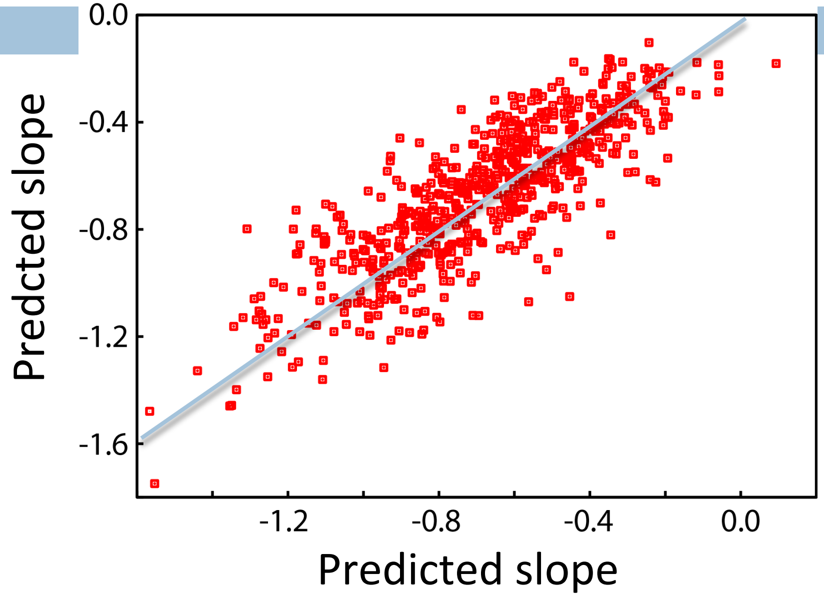
Different solvers predict slow or fast progressors more reliably.

Larger (absolute) errors in case of steep slopes.

Similarity among Predictions



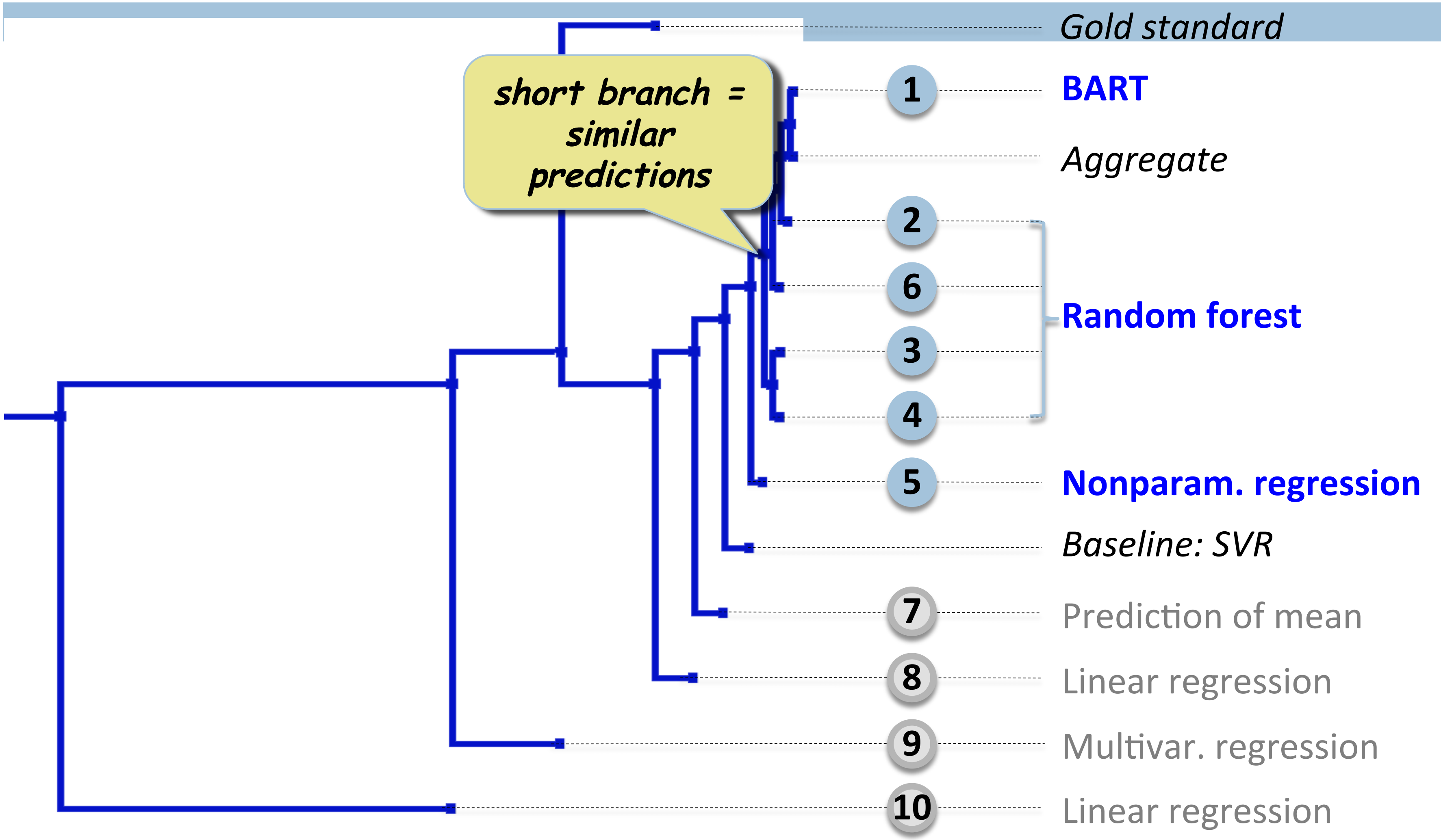
Slopes vs. Predictions



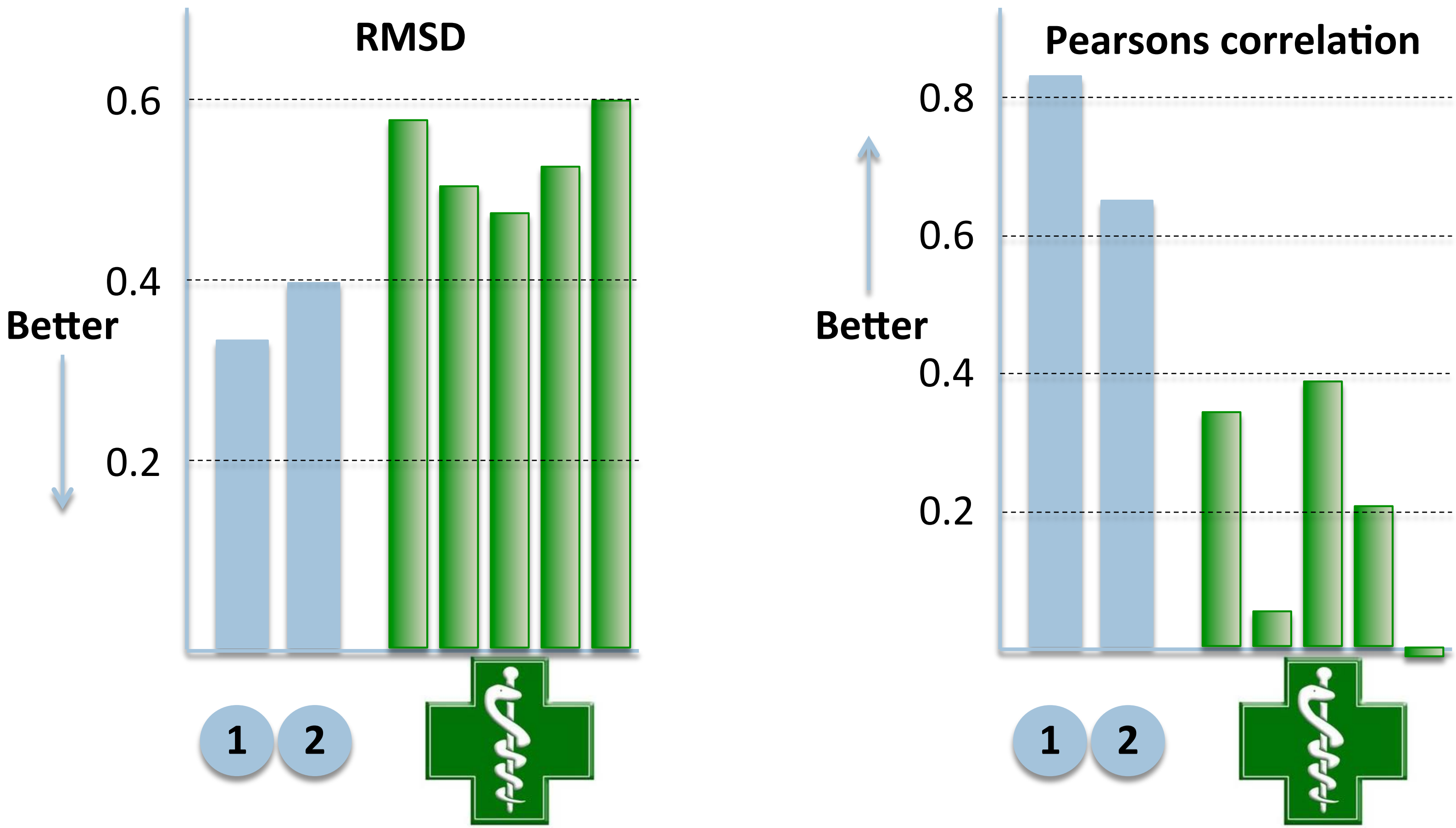
Predictions first vs second

Predictions more correlated to each other than to real slopes: room for improvement?

Similarity among Predictions

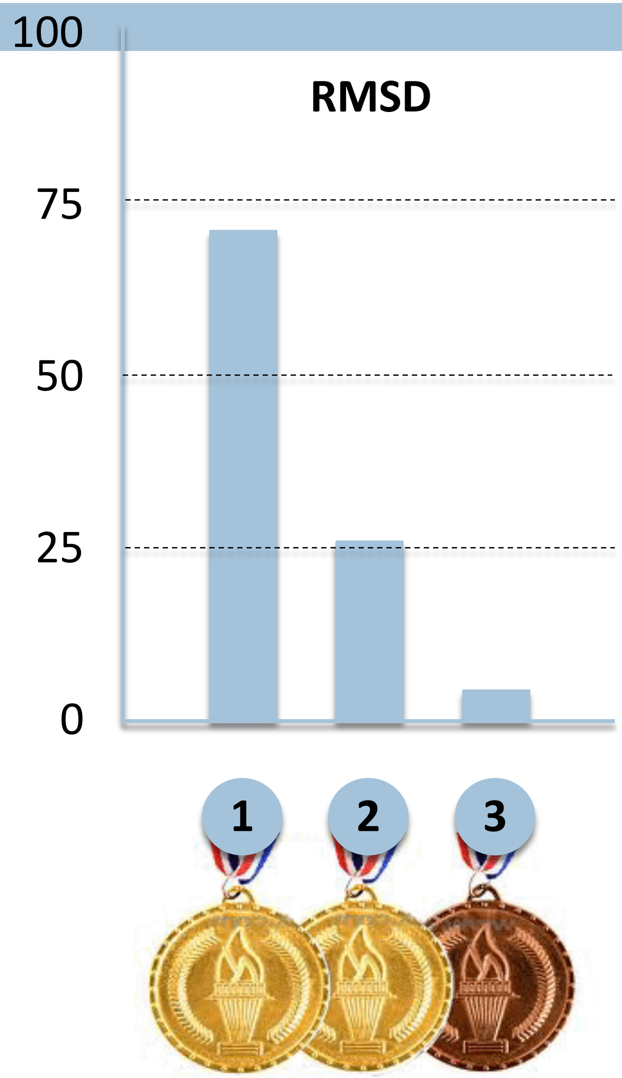


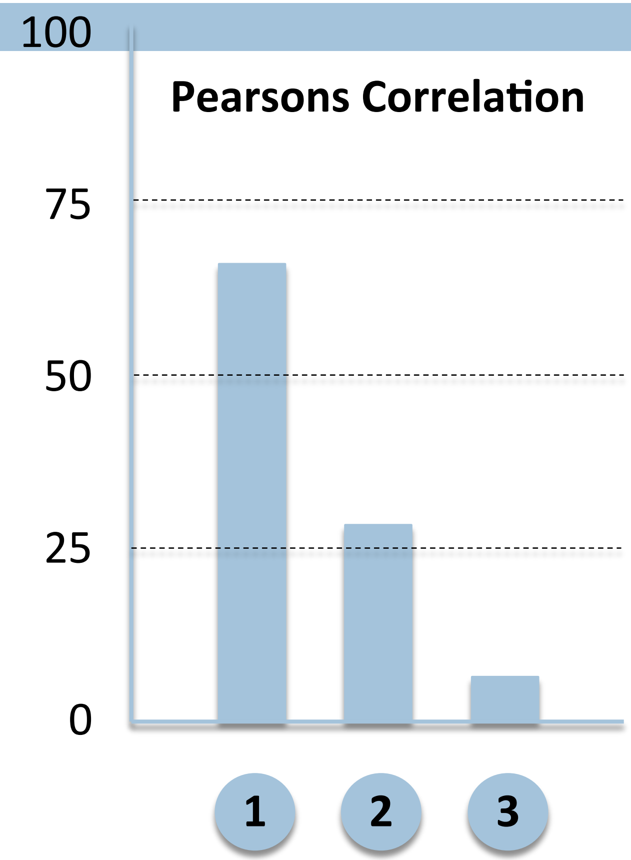
Algorithms vs. Clinicians



Based on 14 patients.

Robustness of Ranking





The Future

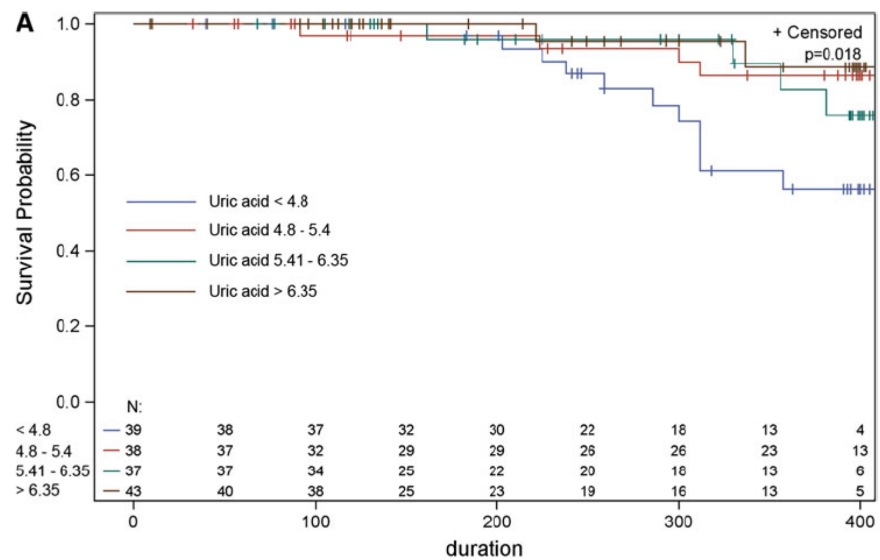
The Future: New ALS Predictors?

Four solvers identify **uric acid** as predictive of progression

- Reported once in the literature but not routinely used

New predictors supported by three or more solvers

- **Pulse**
- **Blood pressure**
- **Creatinine**
- **Basophils**
- **Monocytes**
- **Creatine kinase**



⇒ New lines of inquiry for ALS

Open Question: Better biomarkers based on predictive features?

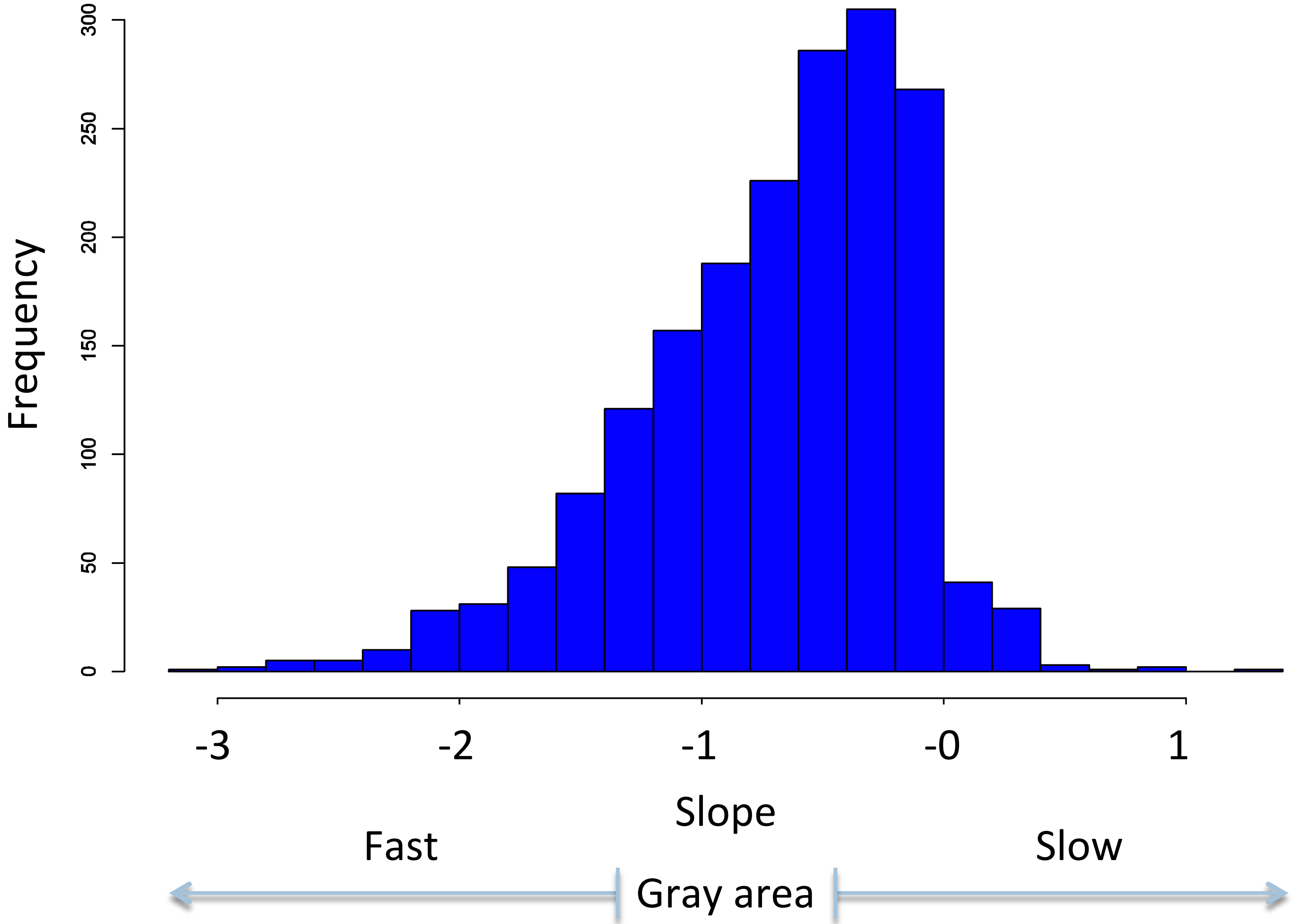
The Future: Clinical Adoption?

- **Grand Challenge:** Introduce algorithms to clinicians, trial managers, and pharmaceutical companies
  - More accurate prognoses for ALS patients
  - Less expensive, more interpretable clinical trials
  - New incentives for ALS drug development

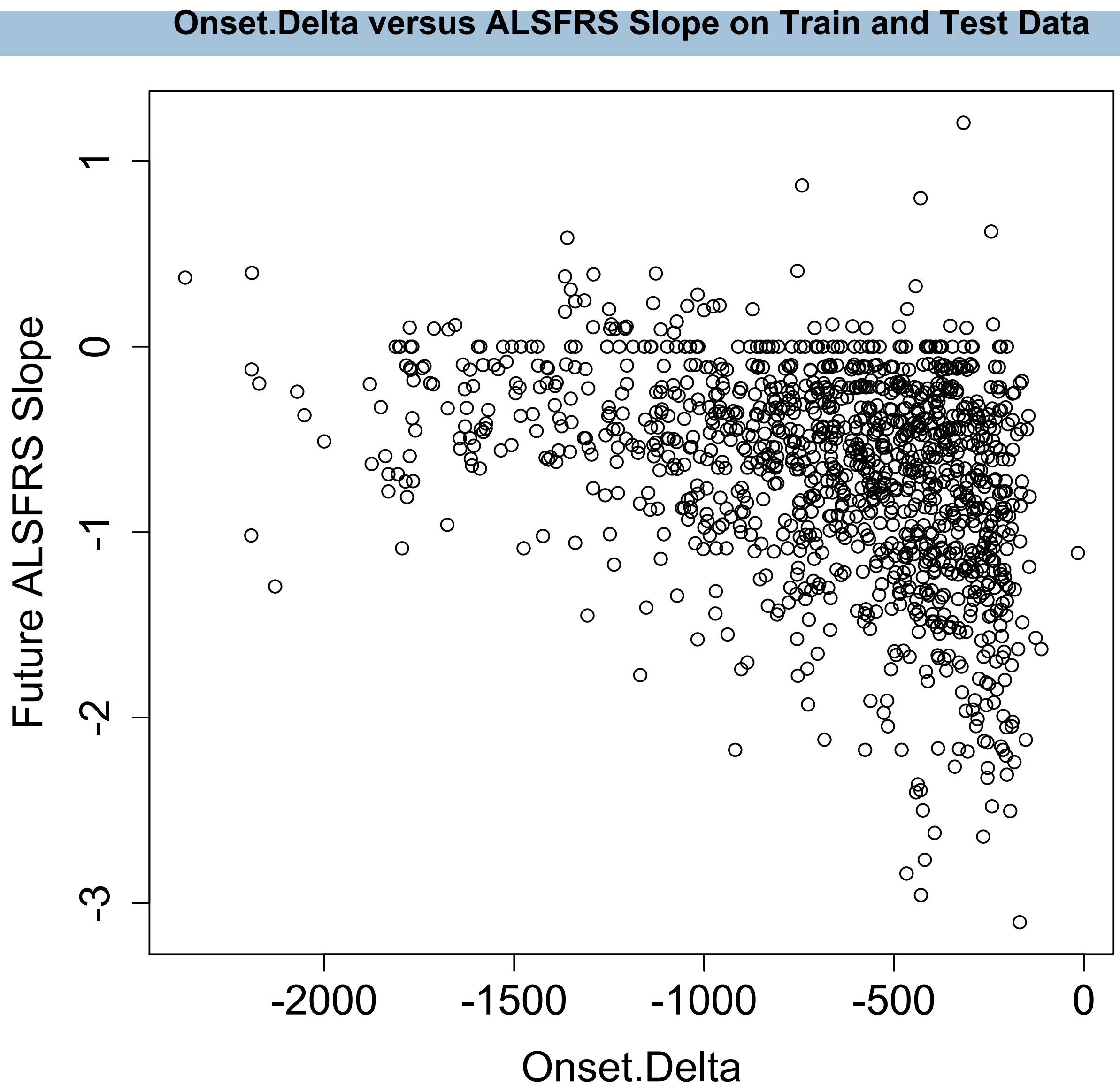
The End

Questions?

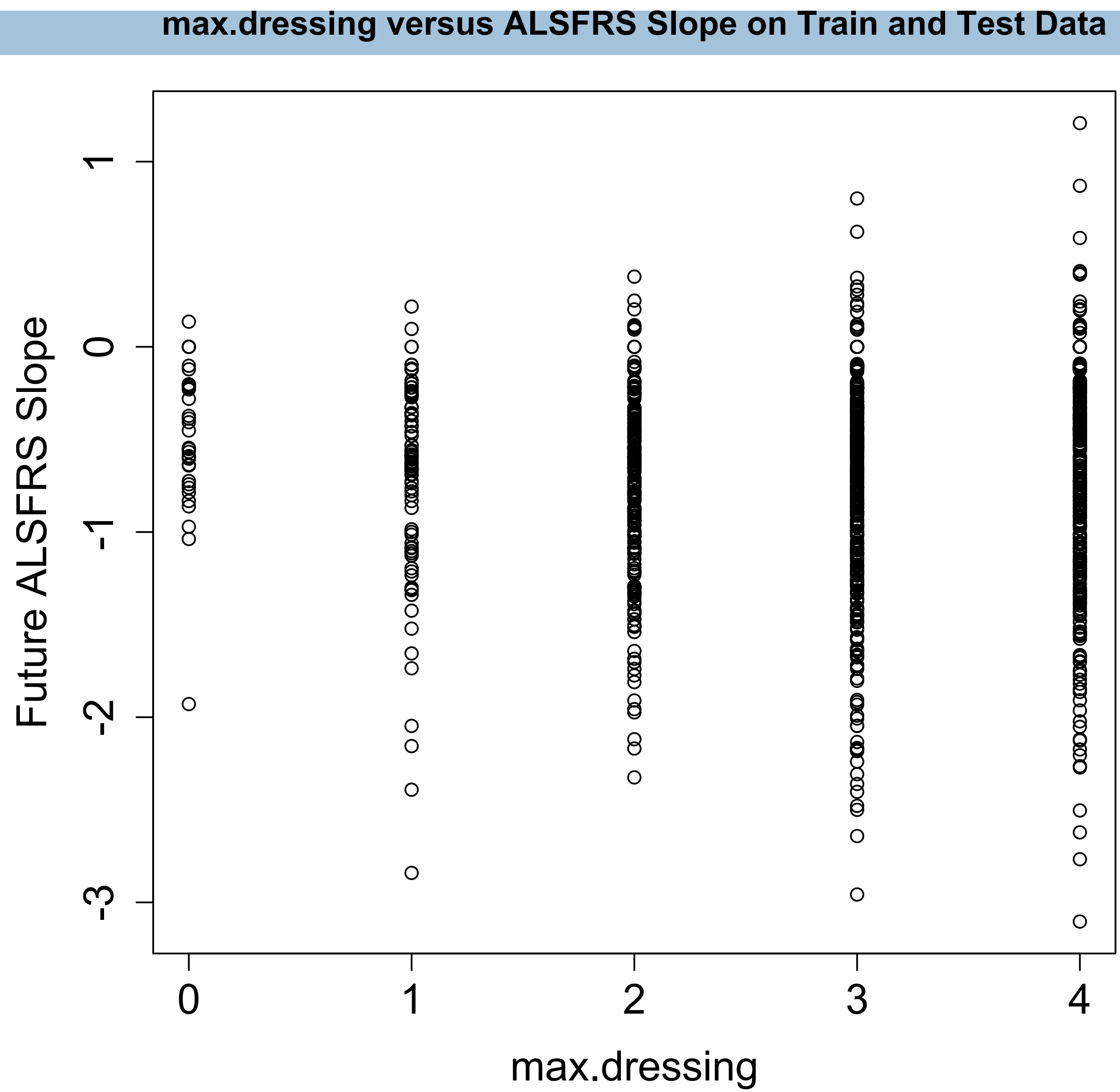
Distribution of ALSFRS Slopes



Onset Delta vs. Target



Max Dressing Score vs. Target



Past ALSFRS Slope vs. Target