Data Science 410
Regularization and Linear Models

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Review

- Introduction to linear regression
  - Linear models are not limited to straight lines
  - Not all parameters are **statistically significant**
  - **Residuals** of linear models, with **least squares loss**, should be **Normally distributed** and **heteroscedastic**
  - Evaluate models with RMSE, adjusted R^2, etc.
  - **High leverage** outliers have significant effect on fit
- Bootstrapping for linear regression
  - Compute bootstrap **distribution of model parameters**
  - Estimate uncertainty in predicted values
- Stepwise regression
  - Forward and backward algorithms **maximize AIC** of selected model
  - Limited by computation and the multiple comparisons problem

Schedule

Part 1	Part 2	Part 3	Part 4
Lesson 1 Data Exploration 1	**Lesson 3** Combinatorics	**Lesson 6** Intro to Regression	**Lesson 9** Näive Bayes
Lesson 2 Data Exploration 2	**Lesson 4** Hypothesis Testing	**Lesson 7** Regularization	**Lesson 10** Basic Text Analysis
Milestone 1 Data Visualization	**Lesson 5** Intro to Bayes	**Lesson 8** Time Series Analysis	**Milestone 4** Independent Project
	Milestone 2 Hypothesis Sim	**Milestone 3** Regression Models	

Reminders!

- Quiz 07 due March 2 – Don’t start until Feb 27
- Discussion 08 must be completed by March 2
- Milestone 03 due March 10
- Milestone 04 due March 13 – no extension possible!
- Assignment 07 due March 3
- Assignment 08 due March 10

It is your responsibility to manage your time for overlapping deadlines!