Administrative

- Project status update
  - Due 11/27 (A week from today by midnight)
  - Take this seriously
  - I want to see some progress
- Quiz
  - Mean: 20.4
  - Median: 19.5
  - Will curve the scores up some (one example, add 10 divide by 35)
- Assignment 4 back soon…

Bias/variance trade-off

We want to fit a polynomial to this, which one should we use?
Bias/variance trade-off

Bias: How well does the model predict the training data?
- high bias – the model doesn't do a good job of predicting the training data (high training set error)
- The model predictions are biased by the model

Variance: How sensitive is the learned model to the training data?
- high variance – changing the training data can drastically change the learned model

High variance OR high bias?

Bias: How well does the model predict the training data?
- high bias – the model doesn't do a good job of predicting the training data (high training set error)
- The model predictions are biased by the model

Variance: How sensitive is the learned model to the training data?
- high variance – changing the training data can drastically change the learned model

High bias

Bias: How well does the model predict the training data?
- high bias – the model doesn't do a good job of predicting the training data (high training set error)
- The model predictions are biased by the model

Variance: How sensitive is the learned model to the training data?
- high variance – changing the training data can drastically change the learned model

High variance

Bias: How well does the model predict the training data?
- high bias – the model doesn't do a good job of predicting the training data (high training set error)
- The model predictions are biased by the model

Variance: How sensitive is the learned model to the training data?
- high variance – changing the training data can drastically change the learned model

High variance
Bias/variance trade-off

What do we want?

Bias: How well does the model predict the training data?
- high bias – the model doesn't do a good job of predicting the training data (high training set error)
- The model predictions are biased by the model

Variance: How sensitive to the training data is the learned model?
- high variance – changing the training data can drastically change the learned model

Bias/variance trade-off

Compromise between bias and variance

Bias: How well does the model predict the training data?
- high bias – the model doesn't do a good job of predicting the training data (high training set error)
- The model predictions are biased by the model

Variance: How sensitive to the training data is the learned model?
- high variance – changing the training data can drastically change the learned model

k-NN vs. Naive Bayes

How do k-NN and NB sit on the variance/bias spectrum?

k-NN has high variance and low bias.
- more complicated model
- can model any boundary
- but very dependent on the training data

NB has low variance and high bias.
- Decision surface has to be linear
- Cannot model all data
- but, less variation based on the training data

Bias vs. variance: Choosing the correct model capacity

Which separating line should we use?
A strong high-bias assumption is linear separability:
- in 2 dimensions, can separate classes by a line
- in higher dimensions, need hyperplanes

Lots of linear classifiers
- Naïve Bayes
- Perceptron
- Rocchio
- Logistic regression
- Support vector machines (with linear kernel)
- Linear regression

Despite this similarity, noticeable performance difference

How might algorithms differ?
Which examples are important?

Another intuition

- If you have to place a fat separator between classes, you have less choices, and so the capacity of the model has been decreased.
Support Vector Machines (SVM)

- SVMs maximize the margin around the separating hyperplane.
- AKA large margin classifiers
- The decision function is fully specified by a subset of training samples, the support vectors.
- Solving SVMs is a quadratic programming problem
- Seen by many as the most successful current text classification method*

Decision trees

- Tree with internal nodes labeled by terms/features
- Branches are labeled by tests on the weight that the term has
  - farm vs. not farm
  - x > 100

*But other discriminative methods often perform very similarly

Decision trees

- Roots are labeled with the class

- Classifier categorizes a document by descending tree following tests to leaf
- The label of the leaf node is then assigned to the document
Decision trees

- Most decision trees are binary trees
- DT make good use of a few high-leverage features
- Linear or non-linear classifier?