Lecture 7: Bias and Variance

Example 1:

Data: + + + + − − − − + + + +

Classifiers: All "thresholds"

Predict + | Predict −

Example 2:

(x_i, y_i), $i = 1, .., n$: training data

$f(x) = y_i$ if $x = y_i$
$= 0$ otherwise.

D: underlying data distribution.

True error of a classifier h is $\Pr_{(X,Y) \sim D} (h(X) \neq Y)$

Concept class: set of classifiers of a certain type.

e.g. concept class of all linear classifiers in d-dimensional space, or concept class of all decision trees.

Bias is the true error of the best classifier in the concept class

eg. best linear separator, best decision tree, etc.

Variance is the extra true error of the trained classifier wrt the best classifier in the concept class. eg. variance is the true error of the classifier you got from the perceptron algorithm − true error of best linear classifier.

In practice, we have both bias and variance.

Classification error comes from Bias + Variance + Noise

$\downarrow$

Random noise; essentially uncertainty in y given x.

When do we have high bias?

High bias when the concept class cannot model the true data distribution well; this doesn't depend on the training data size.

When you have high bias, its called underfitting

When do we have high variance?

High variance when there is a small amount of training data and a very complicated concept class.

When you have high variance, it is called overfitting

Variance decreases with larger training data, and increases with more complicated classifiers.

Examples:

1. Decision trees with k+1 nodes: more complicated than decision trees with k nodes.
2. Linear classifiers: higher the dimension of the feature space, more complex
3. Quadratic classifiers (or higher order polynomial) classifiers are more complex than linear.

High bias $\Rightarrow$ High training and test errors

High variance $\Rightarrow$ Low training error, high test errors.

Bias-variance tradeoff: If we make the concept class more complicated, then, for the same training set size, bias decreases but variance increases. Thus there is a bias-variance tradeoff.