Chapter 2 Overview of Statistical Learning

1. Suppose we observe a quantitative response and p different predictors . Assume there is some relationship between and :

Here f is some fixed but unknown function of , and is a random error term independent of and has mean zero.

1. For any estimate of , we have

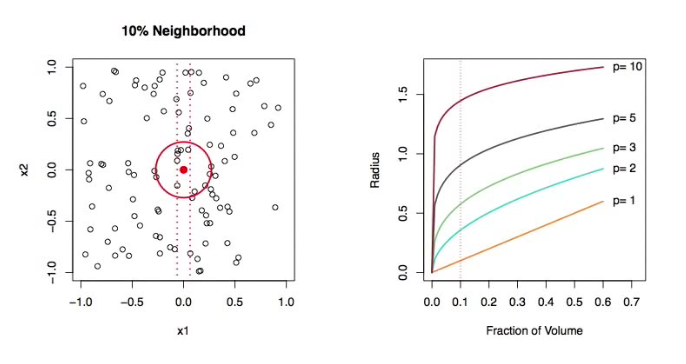
where is reducible error since we can potentially improve , and is irreducible error.

1. How to estimate
   1. Option 1: Estimate .

However, we may have few if any data point at exactly, and thus cannot compute it.

* 1. Option 2: Relax the definition and let , where is some neighborhood of x.

However, we may fall into curse of dimensionality: the nearest neighbors tend to be far away in high dimensions. For example, to include 10% of all data in 1-D uniformly distributed data, the fraction of length is only 10%, and thus radius is 5%; but in 2-D uniformly distributed data, the radius of the circle with area 0.1 is 18%.



* 1. Parametric and structured models, e.g., linear models. Need to consider a few trade-offs: (a) prediction accuracy versus interpretability; (b) good fit versus overfit/underfit.

1. How to assess model accuracy: mean square error on training data and test data
2. Bias-Variance trade-off

For a given value ,

where:

* is the expected test MSE at
* is the variance of if we estimated using a different training dataset
* is the bias between the model and the real-life problem

As a general rule, a more flexible model tends to have lower bias and higher variance.

Chapter 3 Linear Regression

1. How to choose the subset the predictors in the regression model?
   1. Forward selection: begin with null model with only an intercept and no predictors. Each time add a variable that results in the lowest RSS, until some stopping rule is satisfied.
   2. Backward selection: begin with all variables in the model. Each time remove a variable with the largest p-value, until some stopping rule is satisfied.
2. The hierarchy principle for interaction term: if we include an interaction in a model, we should also include the main effects, even if the p-values associated with main effects are not significant.

Rationale: interaction term measures the change of a main effect when another main affect changes. The interaction term would be hard to interpret without main effects in the model.

1. Potential problems of linear regression: (not included in the lectures; see text section 3.3.3; to be added later)
   1. Non-linearity relationship
   2. Correlation of error
   3. Non-constant variance of error terms
   4. Outliers
   5. High-leverage points
   6. Collinearity

Chapter 4 Classification

1. Logistic regression
2. Logistic regression with more than 2 classes

There is a exponential linear function for each class. Only degrees of freedom.