Linear Regression

Miaoyuan (Rachel) Lang

April 23, 2018

Contents

1 Idea

For machine learning purposes, we think of regression as a very simple way to approximate $E(Y|X)$. And we don't need to believe that the linear model assumptions are true.

$$\hat{\mu}(x) = x^T\hat{\beta}$$

Usually, we estimate $\hat{\beta}$ using least square.

2 Least square - Minimizing RSS

RSS (Residual Sum of Squares):

$$RSS(\beta) = ||Y - \hat{Y}||_2^2$$
$$\hat{\beta} = (X^TX)^{-1}X^TY$$
$$\hat{Y} = X(X^TX)^{-1}X^TY = HY$$

3 H is the orthogonal projection onto col{X}

Proof:

1. H is an orthogonal projection.
   H is symmetric ($H = H^T$): $(X(X^TX)^{-1}X^T)^T = X(X^TX)^{-1}X^T$.
   H is idempotent ($H^2 = H$): $[X(X^TX)^{-1}X^T][X(X^TX)^{-1}X^T] = X(X^TX)^{-1}X^T$.
2. H leaves anything in col(X) alone and kills anything orthogonal to col(X).
   Suppose $y \in col(X)$, $y = Xa$, $Hy = X(X^TX)^{-1}X^T(Xa) = Xa = y$.
   Suppose $y \perp col(X)$, $Hy = X(X^TX)^{-1}X^Ty = 0$.

4 Problem - High dimensions

Suppose the linear model is correct, as dimension increases, the prediction error increases (super-)linearly, especially when dimension is bigger than the number of observations (No inverse).

References