Linear Regression: Algorithms and Instabilities

Monday, April 10, 2017 9:32 AM

Goals of Lecture 3

1. Motivate extensions of OLS
2. Matrix decompositions
3. Ridge regression
4. Subset selection : greedy methods

Recall linear regression

$$\underset{n}{\hat{y}} = \underset{n \times p}{X} \ \underset{p}{\hat{\beta}} \qquad \hat{\beta} = (X^T X)^{-1} X^T y$$

$$\hat{y} = \underbrace{X(X^T X)^{-1} X^T}_{H \,:\, \text{hat matrix}} y$$

Fit: solve $X^T X \hat{\beta} = X^T y$

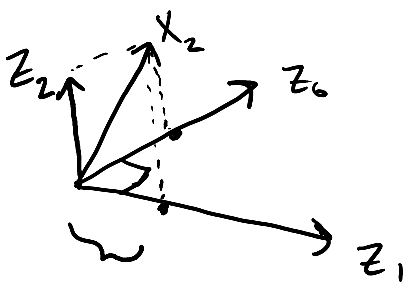
Algorithm 3.1 *Regression by Successive Orthogonalization.*

1. Initialize $\mathbf{z}_0 = \mathbf{x}_0 = \mathbf{1}$.
2. For $j = 1, 2, \ldots, p$

   Regress $\mathbf{x}_j$ on $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_{j-1}$ to produce coefficients $\hat{\gamma}_{\ell j} = \langle \mathbf{z}_\ell, \mathbf{x}_j \rangle / \langle \mathbf{z}_\ell, \mathbf{z}_\ell \rangle$, $\ell = 0, \ldots, j-1$ and residual vector $\mathbf{z}_j = \mathbf{x}_j - \sum_{k=0}^{j-1} \hat{\gamma}_{kj} \mathbf{z}_k$.

3. Regress $\mathbf{y}$ on the residual $\mathbf{z}_p$ to give the estimate $\hat{\beta}_p$.

ESL pg. 54



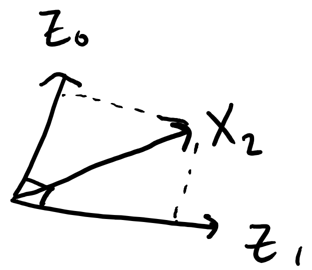
x_j – n dim column of X

$$\frac{\langle z_1, x_2 \rangle}{\langle z_1, z_1 \rangle} = \hat{\gamma}_{12}$$

$$\hat{\beta}_p = \frac{\langle z_p, y \rangle}{\langle z_p, z_p \rangle}$$

!!! Instable if $\langle z_p, z_p \rangle$ is small !!!

Impossible if $\langle z_p, z_p \rangle = 0$



$\Rightarrow \quad z_2 = 0$

linear dependence (always happens if $p > n$)