Linear transformations of covariates in regression:

LS in original variables:

① $\| \underset{\sim}{Y} - X\beta \|_2^2 = \sum_{i=1}^{n} (y_i - x_i^T \beta)^2$,

x_i^T is the ith row of the design matrix X.

X: $(p+1) \times (p+1)$

② using linearly transformed variables: $B = X \cdot A$

B: $n \times (p+1)$

$\| Y - B \cdot b \|_2^2$

① $\| Y - \underbrace{X\beta}_{X \cdot I \cdot \beta = \underbrace{X \cdot A}_{B} \cdot \underbrace{A^{-1} \cdot \beta}_{b}} \|_2^2$

$\Rightarrow$ if $\hat{\beta}$ minimizes ①, then $\hat{b} = A^{-1} \cdot \hat{\beta}$ minimizes ②

if $\hat{b}$ minimizes ②, then $\hat{\beta} = A \cdot \hat{b}$ is the minimizer of ①.

Case 1: recentering covariates

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} = \underbrace{\left(\underset{\sim}{1}, \underset{\sim}{X_1}, \ldots, \underset{\sim}{X_p} \right)}_{\text{columns of } X},$$

$$A_C = \begin{pmatrix} 1 & -\bar{x}_1 & -\bar{x}_2 & \cdots & -\bar{x}_p \\ & 1 & & 0 & \\ & & \ddots & & \\ & 0 & & 1 & \\ & & & & 1 \end{pmatrix}$$; $\bar{x}_j$ is the mean of X_j.

A_C is of full rank.

$$X \cdot A_C[,2] = \underset{\sim}{X_1} - \bar{x}_1 \cdot \underset{\sim}{1}$$

Case 2: (rc) scaling of features.

Idea: divide X_j by its sample std. dev., to get $\tilde{X}_j$ that has unit (sample) variance.

$$A_S = \begin{pmatrix} 1 & & & \\ & 1/\sqrt{s_1^2} & & 0 \\ & & \ddots & \\ & 0 & & 1/\sqrt{s_p^2} \end{pmatrix}$$; s_j^2 is the sample variance for X_j.

case 3: centering and scaling.

$$\tilde{X} = (X \cdot A_C) \cdot A_S .$$

Lasso & ridge regression w/o the intercept.

let $\tilde{y} = y - \bar{y} \cdot \underset{\sim}{1}$

$\bar{y}$: the sample mean of y.

let $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_p)$, where

$$\tilde{X}_j = \left(X_j - \bar{x}_j \cdot \underset{\sim}{1}\right) / \sqrt{s_j^2}$$

equivalent criteria are

$$\| \tilde{y} - \tilde{X}\tilde{\beta} \|_2^2 + \lambda \sum_{j=1}^{p} |\tilde{\beta}_j|^q ,$$

note: no intercept

$q = 1$: lasso

$q = 2$: ridge regression

Singular value decomposition. (SVD)

X be $n \times p$ matrix, $n \geq p$. E.g., X_j's centered and scaled; no intercept.

SVD of X is : $X = U \cdot D \cdot V^T$, such that (U: $n \times p$, D: $p \times p$, V: $p \times p$)

$U^T U = I_p$. $V^T V = I_p$. D is a diagonal matrix.

$D_{11} \geq D_{22} \geq \cdots \geq D_{pp} \geq 0$; the d_j's are called singular values.

(d_1 d_2 d_p)

# of nonzero d_j's is the rank of X.

Often, PCA is defined in terms of SVD (spectral decomp.) of (X'*X).

original LS criterion is

① $\| Y - X\beta \|_2^2$

From svd, $X = U \cdot \underbrace{D \cdot V^T}_{A}$

Alternatively, one can take A = V'

② $\| Y - \underbrace{Ub}_{\sum_{j=1}^{p} u_j b_j} \|_2^2$: no dimension reduction

$\| Y - \sum_{j=1}^{M} u_j b_j \|_2^2$: principal components regression.

Dimension reduction: $M < p$.

We use only a subset of u_j's.