FYS-STK4155
week36 exercises

Yevhenii Volkov

September 10, 2023

Exercise 1

Cost function for Ridge regression

$$C(\beta) = \frac{1}{n}(y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta. \tag{1}$$

We can rewrite it using $w = (y - X\beta) \in \mathcal{R}^n$

$$C(\beta) = \frac{1}{n}w^T w + \lambda\beta^T\beta. \tag{2}$$

Find optimal parameters $\hat{\beta}$ means take derivative of cost function with respect to β and make it equal to zero.

$$\frac{\partial C(\beta)}{\partial\beta} = \frac{2}{n}w^T\frac{\partial w}{\partial\beta} + \lambda\beta^T. \tag{3}$$

Where

$$\frac{\partial w}{\partial\beta} = \frac{\partial(y - X\beta)}{\partial\beta} = -X \tag{4}$$

This result to

$$\frac{\partial C(\beta)}{\partial\beta} = -\frac{2}{n}w^T X + \lambda\beta^T = -\frac{2}{n}(y - X\beta)^T X + \lambda\beta^T = 0. \tag{5}$$

We can take the transpose of this expression

$$\frac{\partial C(\beta)}{\partial\beta^T} = -X^T(y - X\beta) + \lambda\beta = 0. \tag{6}$$

$$\begin{aligned}
-X^T y + X^T X\beta + \lambda I\beta &= 0; \\
X^T y &= X^T X\beta + \lambda I\beta; \\
\hat{\beta} &= (X^T X + \lambda I)^{-1}X^T y.
\end{aligned}$$

Here $I \in \mathcal{R}^{p\times p}$ is an identity matrix. Optimal parameters for Ridge regression

$$\boxed{\hat{\beta}^{Ridge} = (X^T X + \lambda I)^{-1} X^T y}. \tag{7}$$

For Ordinary Least Squares (OLS) the derivation has the same steps, but without λI term.

$$\boxed{\hat{\beta}^{OLS} = (X^T X)^{-1} X^T y} \tag{8}$$

Exercise 2

Now we can use equations (8) and (7) to find expression for prediction vector $\tilde{y}$ using singular value decomposition.

Singular value decomposition

$$X = U\Sigma V^T, \tag{9}$$

where $U \in \mathcal{R}^{n\times n}, V \in \mathcal{R}^{p\times p}$ - orthogonal matrices, $\Sigma \in \mathcal{R}^{n\times p}$ with singular values on the diagonal.

$$\begin{aligned}
\tilde{y}^{OLS} =& X\tilde{\beta}^{OLS} = U\Sigma V^T((U\Sigma V^T)^T U\Sigma V^T)^{-1}(U\Sigma V^T)^T y \\
=& U\Sigma V^T (V\Sigma^T U^T U\Sigma V^T)^{-1} V\Sigma^T U^T y \\
=& U\Sigma V^T (V\Sigma^2 V^T)^{-1} V\Sigma^T U^T y \\
=& U\Sigma V^T (\Sigma^2 V V^T)^{-1} V\Sigma^T U^T y \\
=& U\Sigma V^T V\Sigma^{-2} V^T V\Sigma^T U^T y \\
=& UU^T y = \sum_{j=0}^{p-1} u_j u_j^T y_j.
\end{aligned}$$

On the line 2 we have used that $(ABC)^T = C^T B^T A^T$,
on the line 3 we have used $U^T U = I$,
on the line 4 $V\Sigma^2 V^T = \Sigma^2 V V^T$,
on the line 5 $V^T V = I, \Sigma\Sigma^{-2}\Sigma^T = I$. For Ridge regression

$$\begin{aligned}
\tilde{y}^{Ridg} =& X\tilde{\beta}^{Ridge} = U\Sigma V^T((U\Sigma V^T)^T U\Sigma V^T + \lambda I)^{-1}(U\Sigma V^T)^T y \\
=& U\Sigma V^T (V\Sigma^T U^T U\Sigma V^T + \lambda I)^{-1} V\Sigma^T U^T y \\
=& U\Sigma V^T (V\Sigma^2 V^T + \lambda I)^{-1} V\Sigma^T U^T y \\
=& U\Sigma V^T (\Sigma^2 + \lambda I)^{-1} V\Sigma^T U^T y \\
=& \frac{\Sigma^2}{\Sigma^2 + \lambda I} UU^T y = \sum_{j=0}^{p-1} u_j u_j^T \frac{\sigma_j^2}{\sigma_j^2 + \lambda} y_j.
\end{aligned}$$