WEEK 36

FYS-STK4155 Nadia Ørning

a) Show that $\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y$, I is $p \times p$ ident. matr.

We have already showed that $\frac{\delta (x-As)^T (x-As)}{\delta s} = -2(x-As)^T A$

So for the differentiation of the OLS cost function it is:

$$\frac{\delta C_{OLS}(\beta)}{\delta \beta} = \frac{\delta (y-X\beta)^T (y-X\beta)}{\delta \beta} = -2(y-X\beta)^T X$$

This gives us the optimal $\hat{\beta}_{OLS}$ when $\frac{\delta C_{OLS}}{\delta \beta} = 0$

For the Ridge cost function, all we add is the term $\lambda |\beta|^2$

$\Rightarrow C_{Ridge} = C_{OLS} + \lambda |\beta|^2$

So for differentiation we have:

$$\frac{\delta C_{Ridge}(\beta)}{\delta \beta} = \frac{\delta C_{OLS}(\beta)}{\delta \beta} + \frac{\delta}{\delta \beta}(\lambda |\beta|^2) = -2(y-X\beta)^T X + \frac{\delta}{\delta \beta}\left(\lambda \sum_{i=0}^{p-1} \beta_i^2\right)$$

$$= 2\beta^T X^T X - 2y^T X + \lambda \left(\frac{\delta \sum_{i=0}^{p-1} \beta_i^2}{\delta \beta_0} \dots \frac{\delta \sum_{i=0}^{p-1} \beta_i^2}{\delta \beta_{p-1}} \right)$$

$$= 2\beta^T X^T X - 2y^T X + \lambda (2b_0 \dots 2b_{p-1}) = 2\beta^T X^T X - 2y^T X + 2\lambda \beta^T$$

If the differentiation equal zero, we can solve for $\hat{\beta}$:

$$2\hat{\beta}^T X^T X - 2y^T X + 2\lambda \hat{\beta}^T = 0$$

$$\hat{\beta}^T X^T X + \lambda \hat{\beta}^T = y^T X$$

$$\hat{\beta}^T (X^T X + \lambda I) = y^T X$$

$$\hat{\beta} (X^T X + \lambda I)^T = X^T y$$

$$\hat{\beta} (X^T X + \lambda I) = X^T y$$

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y \quad \square$$

b) Show that $\tilde{y}_{OLS} = X\beta = \sum_{j=0}^{p-1} u_j u_j^T y$

$$\tilde{y}_{OLS} = X\hat{\beta}_{OLS} = X(X^TX)^{-1}X^Ty$$
$$= U\Sigma V^T((U\Sigma V^T)^T U\Sigma V^T)^{-1}(U\Sigma V^T)^T y$$
$$= U\Sigma V^T(V\Sigma^T U^T U\Sigma V^T)^{-1} V U^T \Sigma^T y$$

We use that $V^T = V^{-1}$, $\Sigma^T\Sigma = \begin{pmatrix}\tilde{\Sigma} \\ 0\end{pmatrix}(\tilde{\Sigma}\ 0) = \tilde{\Sigma}^2$, and that

$$\tilde{\Sigma} = \begin{pmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_{p-1}^2 \end{pmatrix}$$

$$\Rightarrow \quad U\Sigma V^T(V\tilde{\Sigma}^2V^T)^{-1} V\Sigma^T U^T y$$
$$= U\Sigma V^T V\tilde{\Sigma}^{-2}V^T V\Sigma^T U^T y = U\Sigma\tilde{\Sigma}^{-2}\Sigma^T U^T y$$

Also $\Sigma\tilde{\Sigma}^{-2}\Sigma^T = I$:

$$UU^Ty = \sum_{j=0}^{p-1} u_j u_j^T y \qquad \square$$

Show for Ridge expression:

$$\tilde{y}_{Ridge} = X\beta_{Ridge} = X(X^TX + \lambda I)^{-1}X^Ty$$
$$= U\Sigma V^T(V\tilde{\Sigma}^2V^T + \lambda I)^{-1}(U\Sigma V^T)^T y$$
$$= U\Sigma V^T(V\tilde{\Sigma}^2V^T + VV^T\lambda I)^{-1} V\Sigma^T U^T y$$
$$= U\Sigma V^T(V(\tilde{\Sigma}^2 + \lambda I)V^T)^{-1} V\Sigma^T U^T y$$
$$= U\Sigma V^T(V^T)^{-1}(\tilde{\Sigma}^2 + \lambda I)^{-1}(V)^{-1} V\Sigma^T U^T y$$
$$= U\Sigma V^T V(\tilde{\Sigma}^2 + \lambda I)^{-1} V^T V\Sigma^T U^T y$$
$$= U\Sigma(\tilde{\Sigma}^2 + \lambda I)^{-1}\Sigma^T U^T y = UU^T \frac{\Sigma\Sigma^T}{\tilde{\Sigma}^2 + \lambda I} y$$

$$= UU^T \frac{\tilde{\Sigma}^2}{\tilde{\Sigma}^2 + \lambda I} y = \sum_{j=0}^{p-1} u_j u_j^T \frac{\sigma_j^2}{\sigma_j^2 + \lambda} y \qquad \square$$