1. Analytical Exercises.

a) Ridge regression: $\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{n} \| \vec{y} - \mathbf{X}\vec{\beta} \|_2^2 + \lambda \|\vec{\beta}\|^2 \right\}$

In vector notation the cost function reads $C(\beta) = (\vec{y} - \mathbf{X}\vec{\beta})^T (\vec{y} - \mathbf{X}\vec{\beta}) + \lambda \vec{\beta}^T \vec{\beta}$

Then it's minimized at $\dfrac{\partial C(\beta)}{\partial \vec{\beta}} = 0$

- $$\frac{\partial (\vec{y} - \mathbf{X}\vec{\beta})^T (\vec{y} - \mathbf{X}\vec{\beta})}{\partial \vec{\beta}} = -2(\vec{y} - \mathbf{X}\vec{\beta})^T \mathbf{X}$$

  Weekly ex. 35.

- $$\vec{\beta}^T\vec{\beta} = \beta_0^2 + \beta_1^2 + \cdots + \beta_i^2 + \cdots + \beta_{n-1}^2$$
  $$\frac{\partial(\vec{\beta}^T\vec{\beta})}{\partial \beta_i} = 2\beta_i$$
  $$\frac{\partial(\vec{\beta}^T\vec{\beta})}{\partial \vec{\beta}} = [2\beta_0 \; 2\beta_1 \; \cdots] = 2\vec{\beta}^T$$

Then
$$\frac{\partial C(\beta)}{\partial \vec{\beta}} = -\cancel{2}(\vec{y} - \mathbf{X}\vec{\beta})^T \mathbf{X} + \cancel{2}\lambda\vec{\beta}^T = 0$$
$$\lambda\vec{\beta}^T = (\vec{y} - \mathbf{X}\vec{\beta})^T \mathbf{X}$$
$$\lambda\vec{\beta} = \mathbf{X}^T(\vec{y} - \mathbf{X}\vec{\beta}) = \mathbf{X}^T\vec{y} - \mathbf{X}^T\mathbf{X}\vec{\beta}$$
$$\lambda\vec{\beta} + \mathbf{X}^T\mathbf{X}\vec{\beta} = \mathbf{X}^T\vec{y}$$
$$(\lambda\mathbb{1} + \mathbf{X}^T\mathbf{X})\vec{\beta} = \mathbf{X}^T\vec{y}$$

Finally: $$\boxed{\vec{\beta}_{Ridge} = (\lambda\mathbb{1} + \mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{y}}$$

If $\mathbf{X} \in \mathbb{R}^{n \times p}$
$\mathbf{X}^T\mathbf{X} \in \mathbb{R}^{p \times p}$
$\Rightarrow \mathbb{1} \in \mathbb{R}^{p \times p}$ identity.

For $\lambda = 0$ we get OLS result: $$\vec{\beta}_{OLS} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\vec{y}$$

b) SVD in OLS Decompose $X \in \mathbb{R}^{n \times p}$

$$\Rightarrow X = U \Sigma V^T \quad \text{with} \quad U \in \mathbb{R}^{n \times n} \quad \Sigma \in \mathbb{R}^{n \times p} \quad V \in \mathbb{R}^{p \times p}$$

Orthogonal
$U U^T = \mathbb{1}$
$V V^T = \mathbb{1}$

$$\Rightarrow X^T = V \Sigma^T U^T$$

Optimal betas: $\vec{\beta} = [X^T X]^{-1} X^T \vec{y}$

Replacing:

$$\vec{\beta} = [V \Sigma^T U^T U \Sigma V^T]^{-1} V \Sigma^T U^T \vec{y}$$

$$= [V \Sigma^T \Sigma V^T]^{-1} V \Sigma^T U^T \vec{y}$$

$$= [\Sigma V^T]^{-1} [V \Sigma^T]^{-1} V \Sigma^T U^T \vec{y}$$

$$= (V^T)^{-1} \Sigma^{-1} (\Sigma^T)^{-1} \underbrace{V^{-1} V}_{= \mathbb{1}} \Sigma^T U^T \vec{y}$$

$$= (V^T)^{-1} \Sigma^{-1} \underbrace{(\Sigma^T)^{-1} \Sigma^T}_{= \mathbb{1}} U^T \vec{y}$$

$$= [V^T]^{-1} \Sigma^{-1} U^T \vec{y}$$

Therefore:

$$\tilde{y}_{OLS} = X \vec{\beta} = U \Sigma \underbrace{V^T \cdot [V^T]^{-1}}_{} \Sigma^{-1} U^T \vec{y}$$

$$= U \Sigma \Sigma^{-1} U^T \vec{y}$$

$$= U U^T \vec{y}$$

Note that $\tilde{y}_{OLS} = U \Sigma V^T \vec{\beta}$

$$\underbrace{\begin{bmatrix} | & | & & | & & | \\ u_1 & u_2 & \cdots & u_p & \cdots & u_n \\ | & | & & | & & | \end{bmatrix}}_{p \qquad\qquad n-p} \begin{bmatrix} \sigma_1 & & 0 \\ 0 & \sigma_2 & \\ & & \ddots \sigma_p \\ 0 & 0 & 0 \end{bmatrix} V^T$$

The last n-p vectors of U are irrelevant because they are multiplied by zeros in the Σ matrix.

Just the first p vectors are relevant. $\Rightarrow U U^T = \sum_{i=1}^{p} \vec{u}_i \vec{u}_i^T$ or $\sum_{i=0}^{p-1} \vec{u}_i \vec{u}_i^T$

Finally:

$$\boxed{\tilde{y}_{OLS} = U U^T \vec{y} = \sum_{i=1}^{p} \vec{u}_i \vec{u}_i^T \vec{y}}$$

Ridge with SVD

Optimal: $\hat{\beta}_{Ridge} = \left[X^T X + \lambda \mathbb{1}\right]^{-1} X^T \vec{y}$

$$X^T X = \left[U \Sigma V^T\right]^T U \Sigma V^T = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$$

Then:
$$\hat{\beta}_{Ridge} = \left[V \Sigma^T \Sigma V^T + \lambda \mathbb{1}\right]^{-1} V \Sigma^T U^T \vec{y}$$
$$= \left[V \Sigma^T \Sigma V^T + \lambda V V^T\right]^{-1} V \Sigma^T U^T \vec{y}$$
$$= \left[V \left(\Sigma^T \Sigma + \lambda \mathbb{1}\right) V^T\right]^{-1} V \Sigma^T U^T \vec{y}$$
$$= \left(V^T\right)^{-1} \left(\Sigma^T \Sigma + \lambda \mathbb{1}\right)^{-1} V^{-1} \cdot V \Sigma^T U^T \vec{y}$$
$$= \left(V^T\right)^{-1} \left(\Sigma^T \Sigma + \lambda \mathbb{1}\right)^{-1} \Sigma^T U^T \vec{y}$$

Finally
$$\tilde{y}_{Ridge} = X \hat{\beta}_{Ridge}$$
$$= U \Sigma V^T \left(V^T\right)^{-1} \left(\Sigma^T \Sigma + \lambda \mathbb{1}\right)^{-1} \Sigma^T U^T \vec{y}$$
$$= U \Sigma \left(\Sigma^T \Sigma + \lambda \mathbb{1}\right)^{-1} \left(U \Sigma\right)^T \vec{y}$$
$$= U \Sigma \left(\Sigma^2 + \lambda \mathbb{1}\right)^{-1} \left(U \Sigma\right)^T \vec{y}$$

$$\Sigma^2 + \lambda \mathbb{1} = \begin{bmatrix} \sigma_1^2 + \lambda & & 0 \\ & \sigma_2^2 + \lambda & \\ 0 & & \ddots \; \sigma_p^2 + \lambda \end{bmatrix} ; \quad \left(\Sigma^2 + \lambda \mathbb{1}\right)^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2 + \lambda} & & 0 \\ & \frac{1}{\sigma_2^2 + \lambda} & \\ 0 & & \ddots \; \frac{1}{\sigma_p^2 + \lambda} \end{bmatrix}$$

$$U \Sigma = \begin{bmatrix} | & | & & | & & | \\ u_1 & u_2 & \cdots & u_p & \cdots & u_n \\ | & | & & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ 0 & \sigma_2 & \ddots \quad \sigma_p \\ 0 & & 0 \end{bmatrix} = \sum_i^p \sigma_i u_i u_i^T \qquad \left(U \Sigma\right)^T = \sum_k^p \sigma_k u_k u_k^T$$

$$\tilde{y}_{Ridge} = \sum_{j=1}^{p} \left(\sigma_j u_j\right) \frac{1}{\sigma_j^2 + \lambda} \left(\sigma_j u_j\right)^T$$

$$\boxed{\tilde{y}_{Ridge} = \sum_{j=1}^{p} u_j \frac{\sigma_j^2}{\sigma_j^2 + \lambda} u_j^T \vec{y}}$$