Exercise 1

(a)

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\bar{\mathbf{x}}\|^2 &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i-\bar{\mathbf{x}})^\top(\mathbf{x}_i-\bar{\mathbf{x}}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}\Big((\mathbf{x}_i-\bar{\mathbf{x}})^\top(\mathbf{x}_i-\bar{\mathbf{x}})\Big) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}\Big((\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top\Big) \\
&= \mathrm{tr}\left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top\right) \\
&= \mathrm{tr}(\mathbf{\Sigma})
\end{aligned}$$

(b)

$$\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i\|^2 = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\bar{\mathbf{x}}\|^2 = \mathrm{tr}(\mathbf{\Sigma}) = \sum_{k=1}^{p}\mathrm{Var}(X_k) = \sum_{k=1}^{p}1 = p$$

(c) Using Lagrangian multipliers, we have that the augmented loss function is

$$\tilde{\mathcal{L}}(\mathbf{W},\lambda) = \mathrm{tr}(\mathbf{W}^\top\mathbf{\Sigma}\mathbf{W}) + \Lambda\,(\mathbf{I}_q-\mathbf{W}^\top\mathbf{W})$$

taking the partial derivative with respect to $\mathbf{W}$ we have that

$$\frac{\partial\tilde{\mathcal{L}}(\mathbf{W},\lambda)}{\partial\mathbf{W}} = 2\mathbf{\Sigma}\mathbf{W} - 2\Lambda\mathbf{W} = 0 \iff \mathbf{\Sigma}\mathbf{W}^\star = \Lambda\mathbf{W}^\star$$

So $\mathbf{W}^\star \in \mathbb{R}^{p\times q}$ is a matrix containing q eigenvectors of $\mathbf{\Sigma}$, but which ones?

Note that if we plug back matrix $\mathbf{W}^\star$ into the loss function, we get

$$\mathcal{L}(\mathbf{W}^\star) = \sum_{i=1}^{q}\lambda_i$$

To minimize $\mathcal{L}$ we should choose the eigenvectors associated to the q-smallest eigenvalues of $\mathbf{\Sigma}$.

Exercise 2

(a) In multiple linear regression we have the model

$$Y = \beta_0 + \beta_1X_1 + \cdots + \beta_pX_p + \varepsilon$$

Note that if we take the expectation from both sites, we get

$$\mathbb{E}[Y] = \beta_0 + \beta_1 \mathbb{E}[X_1] + \cdots + \beta_p \mathbb{E}[X_p] + \mathbb{E}[\varepsilon]$$

Since the predictors and observations have zero-mean, then $\beta_0 = 0$.

(b) The SVD of the data matrix is $\mathbf{X} = \mathbf{UDV}^\top$ so if we plug this into the expression for $\hat{\beta}$ we get

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = (\mathbf{VDU}^\top \mathbf{UDV}^\top)^{-1} \mathbf{VDU}^\top \mathbf{y} = (\mathbf{VD}^2 \mathbf{V}^\top)^{-1} \mathbf{VDU}^\top \mathbf{y} = \mathbf{VD}^{-2} \mathbf{V}^\top \mathbf{VDU}^\top \mathbf{y}$$

so in the end we get $\hat{\beta} = \mathbf{VD}^{-1} \mathbf{U}^\top \mathbf{y}$ and $\hat{\beta}_i = \sum_{k=1}^{p} \frac{\mathbf{u}_i^\top \mathbf{y}}{d_k} \mathbf{v}_{ik}$

Note also that the predictions with the model are $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{UDV}^\top \mathbf{VD}^{-1} \mathbf{U}^\top \mathbf{y} = \mathbf{UU}^\top \mathbf{y}$

(c) Matrix $\mathbf{Z}$ is the projection of the data matrix on its q-top principal components. We have, therefore:

$$\mathbf{Z} = \mathbf{XV}_q = \mathbf{UDV}^\top \mathbf{V}_q = \mathbf{UD} \begin{bmatrix} \mathbf{I}_q & \\ & \mathbf{0}_{p-q} \end{bmatrix} = \mathbf{UD}_q \quad \text{where} \quad \mathbf{D}_q = \begin{bmatrix} d_1 & & & \\ & \ddots & & \\ & & d_q & \\ & & & \mathbf{0}_{p-q} \end{bmatrix}$$

We calculate the coefficients for the new regression model

$$\hat{\gamma} = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{y} = (\mathbf{D}_q \mathbf{U}^\top \mathbf{UD}_q)^{-1} \mathbf{D}_q \mathbf{U}^\top \mathbf{y} = \mathbf{D}_q^{-1} \mathbf{U}^\top \mathbf{y}$$

and if we take it back to the original space, we get

$$\hat{\beta}^{\text{PCR}} = \mathbf{V}_q \hat{\gamma} = \mathbf{V}_q \mathbf{D}_q^{-1} \mathbf{U}^\top \mathbf{y} \quad \text{and} \quad \hat{\beta}_i^{\text{PCR}} = \sum_{k=1}^{q} \frac{\mathbf{u}_i^\top \mathbf{y}}{d_k} \mathbf{v}_{ik}$$

(d) We notice that the parameters for the linear regression obtained with the q-top principal components is a truncated version of the original least squares parameters. We observe that the terms of the sum depending of small singular values have been discarded.