Problem 3

Part (a)

We know that:

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

Goal: Express this in the form $\hat{\beta}_{\lambda}^{RR} = X^T \hat{\alpha}$. Assume:

$$\hat{\beta}_{\lambda}^{RR} = X^T \hat{\alpha}$$

And so

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

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

The matrix $X(X^T X + \lambda I)^{-1} X^T$ is a projection matrix(as discussed in the chapter of LDA and QDA). For $X \in \mathbb{R}^{n \times p}$, where $p > n$, XX^T is invertible because X has full rank. And so we can see that:

$$\hat{\alpha} = (XX^T)^{-1} X(X^T X + \lambda I)^{-1} X^T y$$

Since XX^T is invertible, we can simplify this to:

$$\hat{\alpha} = (XX^T + \lambda I)^{-1} y$$

Thus, the ridge regression solution $\hat{\beta}_{\lambda}^{RR}$ can be written as:

$$\hat{\beta}_{\lambda}^{RR} = X^T \hat{\alpha}$$

with,

$$\hat{\alpha} = (XX^T + \lambda I)^{-1} y$$

Part b

We know that:

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

When $\lambda = 0$, this reduces to:

$$\hat{\beta}_{\lambda=0}^{RR} = (X^T X)^{-1} X^T y$$

Substituting the reduced $\hat{\beta}_{\lambda=0}^{RR}$:

$$X\hat{\beta}_{\lambda=0}^{RR} = X(X^T X)^{-1} X^T y$$

Similar to before notice that $X(X^T X)^{-1} X^T$ is a projection matrix so:

$$X(X^T X)^{-1} X^T y = y$$

Thus, $\hat{\beta}^{RR}_{\lambda=0}$ is a solution to $X\beta = y$.

3. **Show that $\hat{\beta}^{RR}_{\lambda=0}$ is the Minimum L_2-Norm Solution:**

The linear system $X\beta = y$ generally has infinitely many solutions since $p > n$. The solution $\hat{\beta}^{RR}_{\lambda=0}$ minimizes the L_2-norm.

To see why, recall that $\hat{\beta}^{RR}_{\lambda=0} = (X^T X)^{-1} X^T y$ is the unique solution that minimizes the L_2-norm among all possible solutions. This is because it is derived from the least squares criterion, which inherently seeks to minimize $\|\beta\|$.

For any other solution γ to $X\beta = y$:

$$X\gamma = y$$

The difference between γ and $\hat{\beta}^{RR}_{\lambda=0}$ must lie in the null space of X. Therefore, we can write γ as:

$$\gamma = \hat{\beta}^{RR}_{\lambda=0} + v$$

where v is any vector in the null space of X. Since v is in the null space:

$$Xv = 0$$

To minimize $\|\gamma\|$, we must have $v = 0$ because any non-zero v will increase the norm. Hence:

$$\|\gamma\| \geq \|\hat{\beta}^{RR}_{\lambda=0}\|$$

This shows that $\hat{\beta}^{RR}_{\lambda=0}$ is the minimum L_2-norm solution.

Conclusion

Thus, when $\lambda = 0$, $\hat{\beta}^{RR}_{\lambda=0} = (X^T X)^{-1} X^T y$ is a well-defined solution to $X\beta = y$ and is the minimum L_2-norm solution among all solutions to this linear system.