Machine Learnig Homework 5

Problem 1

Let $\hat{\Phi} = \begin{pmatrix} \Phi \\ \sqrt{\lambda} I_{p \times M} \end{pmatrix} \in \mathrm{R}^{(N+p)\times M}$ be the augmented design matrix and $\hat{\mathbf{z}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{0}_p \end{pmatrix} \in \mathrm{R}^{N+p}$ the augmented target vector. Maximum likelihood estimation of parameters $\mathbf{w}$ for linear regression is equivalent to minimizing energy given by $E(\mathbf{w}) = (\mathbf{z} - \Phi^T\mathbf{w})^T(\mathbf{z} - \Phi^T\mathbf{w})$. The minimizer $\mathbf{w}_{MLE} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{z}$ is agnostic of the form and contents of $\mathbf{w}$, $\mathbf{z}$ or Φ. Therefore, we can write the minimizer of the energy resulting from the augmented design matrix and target vector as

$$\begin{aligned}
\hat{\mathbf{w}}_{MLE} &= (\hat{\Phi}^T\hat{\Phi})^{-1}\hat{\Phi}^T\hat{\mathbf{z}} \\
&= \left(\begin{pmatrix} \Phi^T & \sqrt{\lambda} I^T_{p \times M} \end{pmatrix} \begin{pmatrix} \Phi \\ \sqrt{\lambda} I_{p \times M} \end{pmatrix}\right)^{-1} \begin{pmatrix} \Phi^T & \sqrt{\lambda} I^T_{p \times M} \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ \mathbf{0}_p \end{pmatrix} \\
&= \left(\lambda I_{M \times M} + \Phi^T\Phi\right)^{-1} \Phi^T\mathbf{z} \\
&= \mathbf{w}^{ridge}_{MLE}
\end{aligned} \tag{1}$$