Exercise 5

A)

The problem can be written in matrix form as follows:

$$\mathbf{y} = \begin{bmatrix} y(0) & \dots & y(M) \end{bmatrix}^T$$
$$\mathbf{b} = \begin{bmatrix} b(1) & \dots & b(N) \end{bmatrix}^T$$
$$\mathbf{X} = \begin{bmatrix} x_1(0) & \dots & x_N(0) \\ \vdots & \ddots & \vdots \\ x_1(M) & \dots & x_N(M) \end{bmatrix}$$
$$\mathbf{y} = \mathbf{X}\mathbf{b}$$

so that the MSE can be written as

$$\mathrm{J}(\mathbf{b}) = \frac{1}{M}(\mathbf{y} - \mathbf{X}\mathbf{b})^T(\mathbf{y} - \mathbf{X}\mathbf{b})$$

To compute the LS-estimate $\hat{\mathbf{y}}$, we need to find the global minimizer $\hat{\mathbf{b}}$ of J, which can be found by setting its gradient to zero (it's a quadratic form):

$$\begin{aligned} M\nabla \mathrm{J}(\mathbf{b}) &= -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= 2\mathbf{X}^T\mathbf{X}\mathbf{b} - 2\mathbf{X}^T\mathbf{y} \\ \Rightarrow \hat{\mathbf{b}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ \Rightarrow \hat{\mathbf{y}} &= \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \end{aligned}$$

To see that this solution reflects the orthogonality principle:

$$\begin{aligned} \hat{\mathbf{y}}^T(\hat{\mathbf{y}} - \mathbf{y}) &= \hat{\mathbf{y}}^T\hat{\mathbf{y}} - \hat{\mathbf{y}}^T\mathbf{y} \\ &= \mathbf{y}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \\ &= \mathbf{y}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{0} \end{aligned}$$

B)

We want to find a solution $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$ that obeys the orthogonality principle:

$$\begin{aligned} \hat{\mathbf{y}}^T(\hat{\mathbf{y}} - \mathbf{y}) &= \mathbf{0} \\ \hat{\mathbf{b}}^T\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} - \hat{\mathbf{b}}^T\mathbf{X}^T\mathbf{y} &= \mathbf{0} \\ \mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}^T\mathbf{y} &= \mathbf{0} \\ \hat{\mathbf{b}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \end{aligned}$$