Solutions for Exercise Round 1

Exercise 1. (Mean as the Minimum Mean Square Estimator)

Find the optimal point estimate **a** that minimizes the expected value of the loss function

$$C(\boldsymbol{\theta}, \mathbf{a}) = (\boldsymbol{\theta} - \mathbf{a})^\mathsf{T} \, \mathbf{R} \, (\boldsymbol{\theta} - \mathbf{a}), \tag{1}$$

where $\mathbf{R}$ is a positive definite matrix, and the distribution of the parameter is $\boldsymbol{\theta} \sim p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})$.

Proof. The expected value of the loss function:

$$\begin{aligned} f(\mathbf{a}) &:= \int C(\boldsymbol{\theta}, \mathbf{a}) \, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \, \mathrm{d}\boldsymbol{\theta} \\ &= \int (\boldsymbol{\theta} - \mathbf{a})^\mathsf{T} \, \mathbf{R} \, (\boldsymbol{\theta} - \mathbf{a}) \, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \, \mathrm{d}\boldsymbol{\theta}. \end{aligned} \tag{2}$$

Set the derivative to zero:

$$\begin{aligned} \nabla f(\mathbf{a}) &= \int \nabla \left[(\boldsymbol{\theta} - \mathbf{a})^\mathsf{T} \, \mathbf{R} \, (\boldsymbol{\theta} - \mathbf{a}) \right] \, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \, \mathrm{d}\boldsymbol{\theta} \\ &= \int 2 \, \mathbf{R} \, (\mathbf{a} - \boldsymbol{\theta}) \, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \, \mathrm{d}\boldsymbol{\theta} \\ &= 2 \, \mathbf{R} \, \mathbf{a} - 2 \, \mathbf{R} \int \boldsymbol{\theta} \, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \, \mathrm{d}\boldsymbol{\theta} = 0 \end{aligned} \tag{3}$$

which finally gives

$$\mathbf{a} = \int \boldsymbol{\theta} \, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \, \mathrm{d}\boldsymbol{\theta}. \tag{4}$$

□

Exercise 2. (Linear Least Squares Estimation)

Assume that we have obtained T measurement pairs (x_k, y_k) from the linear regression model

$$y_k = \theta_1 \, x_k + \theta_2, \qquad k = 1, 2, \ldots, T. \tag{5}$$

The purpose is now to derive estimates of the parameters θ_1 and θ_2 such that the following error is minimized (least squares estimate):

$$E(\theta_1, \theta_2) = \sum_{k=1}^{T} (y_k - \theta_1 \, x_k - \theta_2)^2. \tag{6}$$

(a) Define $\mathbf{y} = (y_1 \ \dots \ y_T)^\mathsf{T}$ and $\boldsymbol{\theta} = (\theta_1 \ \theta_2)^\mathsf{T}$. Show that the set of Equations (5) can be written in matrix form as

$$\mathbf{y} = \mathbf{X}\,\boldsymbol{\theta},$$

with a suitably defined matrix $\mathbf{X}$.

Proof. We write $\mathbf{X} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_T & 1 \end{pmatrix}$ $\square$

(b) Write the error function in Equation (6) in matrix form in terms of $\mathbf{y}$, $\mathbf{X}$, and $\boldsymbol{\theta}$.

Proof. $E(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\,\boldsymbol{\theta})^\top(\mathbf{y} - \mathbf{X}\,\boldsymbol{\theta})$ $\square$

(c) Compute the gradient of the matrix form error function, and solve the least squares estimate of the parameter $\boldsymbol{\theta}$ by finding the point where the gradient is zero.

Proof. $\nabla E(\boldsymbol{\theta}) = -2\mathbf{X}^\top\mathbf{y} + 2\mathbf{X}^\top\mathbf{X}\,\boldsymbol{\theta}$, so that $\nabla E(\boldsymbol{\theta}) = 0 \iff \boldsymbol{\theta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$. $\square$