FECD-B, 2026: Homework 01

Renato Assunção

Easy (Concept Check)

1. Explain the fundamental difference between $g(\boldsymbol{x})$ and $g(\boldsymbol{X})$ in the context of predicting a random variable Y.

2. What specific function $g(\boldsymbol{x})$ provides the solution to the optimization problem of minimizing the Mean Squared Error, $MSE(g) = \mathbb{E}[(Y - g(\boldsymbol{X}))^2]$?

3. Suppose the conditional expectation $\mathbb{E}[Y|X = \boldsymbol{x}]$ is estimated directly from a dataset using the empirical arithmetic average of Y values in the examples for which $X = \boldsymbol{x}$. Why does having a large number of features create a practical difficulty (often called the "curse of dimensionality")?

4. In the context of the linear model $(Y|\boldsymbol{X} = \boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$, what is the practical interpretation of the coefficient β_j?

Medium (Application & Calculation)

5. Let Y be a random variable with finite expectation $\mu = \mathbb{E}[Y]$. Let $\varepsilon = Y - \mathbb{E}[Y]$ and consequently $Y = \mathbb{E}[Y] + \varepsilon$. Answer the following:
   (a) Is ε a random variable?
   (b) Is $\mathbb{E}[Y]$ a random variable or a constant?
   (c) Show that $\mathbb{E}(\varepsilon) = 0$.

6. **Distinguishing $g(\boldsymbol{x})$ from $g(\boldsymbol{X})$.** Let $g(x_1, x_2) = 2 + 3x_1 - x_2$. Assume the random vector $\boldsymbol{X} = (X_1, X_2)$ can take only three possible values: $(1, 0)$, $(2, 1)$, and $(0, -1)$ with probabilities 0.2, 0.5, 0.3, respectively.
   (a) Compute the three scalar values $g(1, 0)$, $g(2, 1)$, and $g(0, -1)$.
   (b) List the possible values of the *random variable* $g(\boldsymbol{X})$ and the list of associated probabilities.
   (c) Compute $\mathbb{E}[g(\boldsymbol{X})]$.

7. **Best constant predictor.** Consider the restricted class of predictors $g(\boldsymbol{X}) = c$, where c is a constant. That is, we have no additional information represented by features $\boldsymbol{X}$, and we must predict the value of the random variable Y with a single constant value c. The prediction error is the random variable $\epsilon = Y - c$. Using a given c, we want to know how much we can expect in terms of the prediction error ϵ. Take the expected (squared) prediction error $\mathbb{E}[\epsilon^2] = \mathbb{E}[(Y - c)^2]$. Show that the value of c that minimizes this expected (squared) prediction error is $c = \mathbb{E}(Y)$.

8. Using the iterated expectation formula, write down the mathematical steps to show that minimizing the overall $MSE(g)$ is equivalent to minimizing the inner expectation $h(\boldsymbol{x}) = E_{Y|X}[(Y - g(\boldsymbol{x}))^2 | X = \boldsymbol{x}]$ for each fixed value of $\boldsymbol{x}$.

9. Prove that the function $h(\boldsymbol{x}) = E[(Y - g(\boldsymbol{x}))^2 | X = \boldsymbol{x}]$ is minimized exactly when $g(\boldsymbol{x}) = \mu(\boldsymbol{x})$, where $\mu(\boldsymbol{x}) = E[Y|X = \boldsymbol{x}]$. (Hint: start your proof by adding and subtracting $\mu(\boldsymbol{x})$ inside the squared term).

10. Given the true surface $\mu(\boldsymbol{x}) = 1.0 + 0.75x + 0.22\sin((x - 0.3)^2) + 0.04\exp(-(x - 0.3)^2)$, calculate the first-order Taylor approximation (the tangent line) around the mean $\bar{x} = 0.0$. Plot the function and its approximating tangent line.

11. Using the same function $\mu(\boldsymbol{x}) = 1.0 + 0.75x + 0.22\sin((x-0.3)^2) + 0.04\exp(-(x-0.3)^2)$, do the following:

    (a) Compute $\mu''(\boldsymbol{x})$ and $\mu''(0)$.

    (b) Write the second-order Taylor approximation around $\bar{x} = 0$.

    (c) Plot and visually compare the first-order and second-order approximations with the original function $\mu(\boldsymbol{x})$. Which one gives a better local approximation near $x = 0$, and why?

12. For the multivariable function $g(\boldsymbol{x}) = \frac{e^{-3x_1^2}}{1+2x_1x_2+x_2^2}$, compute the gradient vector $\nabla g(\boldsymbol{x})$ evaluated at the point $p = (1,1)$, and use it to write the first-order Taylor expansion (tangent plane) around this point.

Hard (Derivation & Advanced Application)

13. To be completed in the future