CIS3333: Mathematics of Machine Learning Fall 2025

Homework 2

Due: September 17, 2025, 11:59 PM ET

Submission Instructions: Submit a single PDF (clear scans or photos compiled) to Gradescope and assign pages for each problem. Show key steps and justify answers.

Collaboration & AI Policy: You may discuss approaches with classmates, but write up your own solutions and list collaborators. If you use computational tools (including LLMs) for checking, cite them and ensure the reasoning is your own.

Problem: Statistics of a Linear Predictor (24 points)

In this problem, you will get more familiar with random vectors by working through a 2D case. The goal is to derive the mean and variance of a linear predictor, $y = \mathbf{w}^T\mathbf{x} + b$, which is a cornerstone of machine learning. Let our feature vector be $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. The linear predictor is a scalar value y given by:

$$y = \mathbf{w}^\top \mathbf{x} + b$$

where $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, and b are constant (not random).

Part 1: Mean of the Predictor (4 points)

Let the mean of $\mathbf{x}$ be $\mu = \mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mathbb{E}[x_1] \\ \mathbb{E}[x_2] \end{bmatrix}$.

1.1 (4 points) Using the linearity of expectation for scalar random variables ($\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$), show that the expected value of the predictor is:

$$\mathbb{E}[y] = w_1\mu_1 + w_2\mu_2 + b = \mathbf{w}^T\mu + b.$$

Part 2: Building Blocks for Variance (8 points)

To find the variance of y, we first need a few properties. Recall that the covariance matrix of our 2D vector $\mathbf{x}$ is a 2×2 matrix defined as:

$$\mathbf{\Sigma} = \mathbb{V}[\mathbf{x}] = \mathbb{E}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T]$$

2.1 (5 points) Show that $\mathbf{\Sigma} = \begin{bmatrix} \mathbb{V}[x_1] & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_2, x_1) & \mathbb{V}[x_2] \end{bmatrix}$.

2.2 (3 points) Using the result from the previous part, show that $\mathbf{\Sigma} = \mathbb{E}[\mathbf{x}\mathbf{x}^T] - \mu\mu^T$.

Part 3: Variance of the Predictor (8 points)

Now we will derive the variance of the predictor in two steps.

3.1 (4 points) Recall that the variance of a sum of two random variables is $\mathbb{V}[A+B] = \mathbb{V}[A] + \mathbb{V}[B] + 2\text{Cov}(A, B)$. Use this property to show that the variance of our linear predictor is:

$$\mathbb{V}[y] = w_1^2\mathbb{V}[x_1] + w_2^2\mathbb{V}[x_2] + 2w_1w_2\text{Cov}(x_1, x_2).$$

3.2 (4 points) Show that the scalar expression above is equivalent to the compact matrix form $\mathbf{w}^T\mathbf{\Sigma}\mathbf{w}$.

A Note on Higher Dimensions. The compact matrix formulas you've derived, $\mathbb{E}[y] = \mathbf{w}^T\boldsymbol{\mu}+b$, $\Sigma = \mathbb{E}[\mathbf{x}\mathbf{x}^T] - \mu\mu^T$ and $\mathbb{V}[y] = \mathbf{w}^T\mathbf{\Sigma}\mathbf{w}$, are not limited to two dimensions. They hold true for any D-dimensional feature vector $\mathbf{x}$. Proving this generalization is a great optional exercise to test your understanding of linear algebra with random vectors.

Part 4: Application (4 points)

Let's apply these formulas. Suppose a student's final grade, y, is determined by their scores on a midterm (x_1) and a final exam (x_2). The statistics for the class are:

- Mean scores: $\mathbb{E}[x_1] = 70$, $\mathbb{E}[x_2] = 80$.
- Variances: $\mathbb{V}[x_1] = 16$, $\mathbb{V}[x_2] = 25$.
- Covariance: $\text{Cov}(x_1, x_2) = 9$.

The final grade is computed with the formula $y = 0.5x_1 + 0.5x_2$.

4.1 (4 points) Using your results from the previous parts, calculate the mean final grade $\mathbb{E}[y]$ and the variance of the final grade $\mathbb{V}[y]$.