ECE 411 Homework 5 (Fall 2024)
Instructor: Dr. Chau-Wai Wong
Material Covered: Statistical Learning Basics

Problem 1 (20 points) [Conditional Expectation, Variance Operator]

a) Given the joint PMF for random variables X and Y in the table, compute the following

Table 1: Joint PMF, $p_{X,Y}(x,y)$

Y \ X	1	2	3
0	0.3	0.1	0.3
1	0.1	0.1	0.1

quantities and tabulate your results: $p_X(x)$, $p_Y(y)$, $p_{X|Y}(x|y)$, $p_{Y|X}(y|x)$, $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathbb{E}[X|Y=y]$, $\mathbb{E}[Y|X=x]$. (Intermediate steps for calculating expectations must be shown to receive full points.) Explain the difference between $\mathbb{E}[X|Y=y]$ and $\mathbb{E}[X|Y]$.

b) Prove the following formulas:

$$\mathrm{Var}(aX+b) = a^2\ \mathrm{Var}(X), \tag{1a}$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\ \mathrm{Cov}(X,Y), \tag{1b}$$

$$\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y), \text{ when } X \text{ and } Y \text{ are uncorrelated.} \tag{1c}$$

$$\mathrm{Var}\left(\sum_i a_i X_i\right) = \sum_i a_i^2\ \mathrm{Var}(X_i),\ X_i\text{'s uncorrelated.} \tag{1d}$$

You may find the following equations useful: i) the shortcut formulas for variance, $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$; and ii) the covariance, $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. Answer the following questions:

- Why does b not appear on the right-hand side of (1a)?
- How does the variance of sum of two random variables compare to the sum of the variance of individual variables when the variables are negatively/anti-correlated? Can you give an extreme example?
- Why is it a plus sign rather than a minus sign on the right-hand side of (1c)?

Problem 2 (20 points) [Optimality of Mean Operators]

a) We are given two variables X and Y that are not independent. Hence, we may use one to estimate the other. Find the best deterministic function $g(\cdot)$ such that it minimizes the expected squared error between Y and $g(X)$ conditioned on $X=x$. You may find a change of variable using θ in the place of $g(x)$ helpful. Pay attention to write clearly the upper case X and the lower case x in your submission.

b) *Arithmetic average*, or the *sample mean* in a statistical context, is commonly used in everyday life for making quantitative description. We examine a statistical interpretation for the arithmetic average below. A person weighs μ lb. He tried multiple scales in a supermarket and recorded the reading from each scale, denoted by Y_i for the ith scale. We may create a linear model as follows to relate the true weight μ and the measurement Y_i:

$$Y_i = \mu + e_i, \quad i = 1, \ldots, N,$$

where e_i is the measurement error of the ith scale. Use the mean-square criterion $J(\mu) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \mu)^2$ to find the closed-form expression for the best estimator for μ. The expression should contain $\{Y_i\}_{i=1}^{N}$ only, and should not contain such symbols as μ or e_i as they were not available when readings were recorded. Does the expression make intuitive sense?

Problem 3 (20 points) [Alternative Neighbor Averaging Method for Simulated Data]

a) Given a regression function $f(x) = x^2 + 2x + 1$ and a linear model $Y = f(X) + e$, where $e \sim \mathcal{N}(0, 1)$ and $X \sim \text{Uniform}(-1, 1)$, generate 50 pairs of (x_i, y_i) and graph them using black circles. Also, plot the regression function using a black solid curve.

b) We use a method similar to the nearest neighbor averaging to estimate the regression function. We use a neighborhood of fixed radius $\delta = 0.1$. The estimated regression function takes the following form:

$$\hat{f}(x) = \frac{1}{|I(x)|}\sum_{i \in I(x)} y_i, \quad I(x) = \{i : |x - x_i| \le \delta\}, \tag{2}$$

where $I(x)$ is the set of indices of x_i such that they are within δ in terms of distance from x, and $|I(x)|$ is the number of elements of set $I(x)$. For example, when $x = 0.9$ and $\delta = 0.1$, you first need to find all points that are within the range of [0.8, 1.0] in the x-direction, and then take the average of their values in the y-direction to obtain $\hat{f}(0.9)$. You may want to calculate $\hat{f}(\cdot)$ for all $x \in [-0.9, 0.9]$ with a step size of 0.01. If there is not a single point within the current neighborhood, use the $\hat{f}$ from the previous step as that for the current step. Draw the estimated regression function using a red solid curve in the same plot of a).

c) (Bonus, 5 points) Vary the neighborhood radius δ, how does the shape of the estimated regression function change?

Problem 4 (20 points) [Curse of Dimensionality] Read the first paragraph of Ex. 2.4 in the Exercises (last) section of Chapter 2 of *The Elements of Statistical Learning* (12th Printing, 2017). Note that we may also write $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$, where $X_k \sim \mathcal{N}(0, 1)$ for $k = 1, \ldots, p$. Use a programming language of your choice. To get started, set $p = 10$. Note that in this problem, all vectors are column vectors.

a) Write a computer program to randomly draw/generate $N = 100$ vectors from the template random vector $\mathbf{X}$, namely, $\{\mathbf{x}^{(i)},\ i = 1, \ldots, N\}$. Note that each vector should contain p normally distributed random numbers. Plot all vectors as points in a 3-D space consisting of the first, second, the last coordinates.

b) Calculate the coordinate value of each point after being projected onto a fixed direction specified by $\mathbf{a} = \mathbf{x}_0/\|\mathbf{x}_0\|$, namely, $z^{(i)} = \mathbf{a}^T\mathbf{x}^{(i)}$. Here, $\mathbf{x}_0$ is an arbitrary nonzero vector of length p, "T" is the transpose operation, and $z^{(i)} \in \mathbb{R}$. What are the sample mean and sample variance of the projected coordinates $\{z^{(i)},\ i = 1, \ldots, N\}$?

c) Repeat a) and b) for $p \in \{1, 2, \ldots, 80\}$. You may want to use a `for` loop to achieve this. Optionally, put your code for parts a) and b) into a function to make your code easier to read. Plot the sample variance of the projected coordinates as a function of p.

d) Calculate the squared distance of each point to the origin, namely, $d_i^2 = \|\mathbf{x}^{(i)}\|^2$. What is the sample mean of $\{d_i^2,\ i = 1, \ldots, N\}$? Plot the sample mean of the squared distance as a function of p in the same plot of c). Limit the range of y-axis between 0 and 80. For $p = 5$, inspect the values of any five d_i^2's. Do the results in b) and c) match with the conclusion drawn in the third paragraph of *ESLII-2.4*?

e) Prove that $\text{Var}(Z) = 1$ where $Z = \mathbf{a}^T\mathbf{X}$, and $\mathbb{E}[D^2] = p$ where $D = \|\mathbf{X}\|$. Are the theoretical results in this part consistent with the simulated results obtained in c) and d)?