ISE - 364/464: Introduction to Machine Learning
Homework Assignment 2

The goal of this assignment is to provide a series of problems that strengthen skills in analytical judgement, probability, multivariate distributions, hyperplanes, and analysis.

Grading: This assignment is due on Coursesite by E.O.D. 10/04/2024. All problems are worth the same number of points. If a problem has multiple parts, each of those parts will be worth equal amounts and will sum to the total number of points of the original problem (Example: If each problem is worth a single point, and problem 1 has 4 parts, each part will be worth 1/4th of a point). ISE - 364 students are only required to answer problems 1 through 4; however, you are allowed to answer the 5th graduate-level question (if done so correctly, you will receive extra credit in the amount that the 5th problem will be worth for the ISE - 464 students). ISE - 464 students are required to answer all 5 problems.

Submitting: Only electronic submissions on Coursesite are accepted.

1 Problems

1. (**Analytical judgement**)

   a) What do probability mass functions compute? What do probability density functions compute? Explain why they are different.

   b) Explain what Simpson's Paradox is. How is it similar to *The Problem of Induction.* *(Hint: We talked about the problem of induction during the first set of lecture slides; however, you may look up material to get a better understanding of what it is)*

2. (**Probability**) Consider a random experiment (we will call this $e = 1$ for experiment 1) that results in three outcomes: x, y, and z.

   a) Assuming that each outcome is equally likely, fully define the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for this random experiment. That is, write the sample space Ω, the event space $\mathcal{F}$, and the probability measure $\mathbb{P}$ (the probabilities associated with every element in the sample space) for this experiment.

   b) We wish to define X as a random variable describing this experiment. Recall that random variables are mappings from the sample space Ω to a target space $\mathcal{T} \subseteq \mathbb{R}$, i.e., $X : \Omega \to \mathcal{T}$. Fully define the "new" probability space $(\mathcal{T}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ corresponding to the random variable X for this experiment. That is, write the target space $\mathcal{T}$, the new event space $\tilde{\mathcal{F}}$, and the new probability measure $\tilde{\mathbb{P}}$ (the probabilities associated

with every element in the target space) for this experiment.
(Hint: For discrete random variables, target space mappings typically start at 0 and increment by whole numbers.)

c) What is the probability of observing a w from this experiment?

d) What is the probability of observing an x or a y from this experiment (written using the random variable and set theory notation)?

e) Now, consider the new random experiment (we will call this $e = 2$ for experiment 2) of performing the original random experiment $e = 1$ twice. Experiment $e = 2$ can be defined with its own probability space $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Write the new sample space Ω_2 of this random experiment. What will the cardinality (the number of elements in a set) of the event space $\mathcal{F}_2$ be?
(Hint: Recall that the event space is equivalent to the power set of the sample space. As such, you can look up the equation for what the cardinality of power sets is.)

f) Let S denote a random variable of experiment $e = 2$. Define the target space $\mathcal{T}_2$ of this experiment. Using this newly defined random variable, what is the probability of observing a z and a z from two consecutive $e = 1$ experiments?

3. (**Multivariate distributions and covariance matrices**) For the following questions, let the signs "$+$" $= p > 0$, "$-$" $= n < 0$, and "0" $= 0$ denote some positive constant ("$+$"), some negative constant ("$-$"), or 0. These three symbols will be used in the following problems to define the "sign structure".

a) Consider two Gaussian random variables X_1 and X_2 which are both independent and identically distributed (i.i.d.) with means $\mu_1 = \mu_2 = +$, respectively. Draw an illustration of a joint probability distribution that satisfies the conditions of these random variables in Cartesian coordinates. Write the sign structure of the 2×2 covariance matrix Σ (simply write either a "$+$", "$-$", or 0 in each entry).

b) Consider two Gaussian random variables X_1 and X_2 which are both identically distributed with means $\mu_1 = \mu_2 = -$, respectively, but which have a positive linear relationship. Draw an illustration of a joint probability distribution that satisfies the conditions of these random variables in Cartesian coordinates. Write the sign structure of the 2×2 covariance matrix Σ.

c) Consider two Gaussian random variables X_1 and X_2 which are neither independent nor identically distributed with means $\mu_1 = +$ and $\mu_2 = -$, respectively, but which have a negative linear relationship. Draw an illustration of a joint probability distribution that satisfies the conditions of these random variables in Cartesian coordinates. Write the sign structure of the 2×2 covariance matrix Σ.

4. (**Hyperplanes and projections**) For the following questions, let the signs "$+$" $= p > 0$, "$-$" $= n < 0$, and "0" $= 0$ denote some positive constant ("$+$"), some negative constant ("$-$"), or 0. These three symbols will be used in the following problems to define the "sign structure".

a) In the four following problems you will be given the sign structure of normal vectors a and scalars b. Draw the corresponding hyperplanes defined by each pair (a, b) and shade the side of the plane that satisfies $a^\top x \geq b$.

i) $a = \begin{bmatrix} + \\ + \end{bmatrix}$ and $b \geq 0$.

ii) $a = \begin{bmatrix} - \\ - \end{bmatrix}$ and $b \leq 0$.

iii) $a = \begin{bmatrix} + \\ - \end{bmatrix}$ and $b \geq 0$.

iv) $a = \begin{bmatrix} - \\ + \end{bmatrix}$ and $b = 0$.

b) Consider the unit vector $z = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Compute the vector b that z is mapped to under the linear transformation $A = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$ (i.e., what is Az)? Draw both vectors z and b in Cartesian coordinates.

c) Notice that the matrix-vector product Ax is a linear system defined by the two equations $a_1^\top x = b_1$ and $a_2^\top x = b_2$, where $a_1^\top = [2, 1]$ is the top row of the matrix A and $a_2^\top = [2, 3]$ is the bottom row, and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is a vector of variables. Let b_1 and b_2 correspond to the two elements of the vector that z was mapped to under the transformation A in problem 4b. Draw both of these hyperplanes.
(Hint: To draw the planes, you will need to know what their intersection is, which can be computed by solving the linear system for x.)

d) Given a vector $y = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$, compute its projection to the nearest of the two hyperplanes.

5. (ISE-464 Graduate Students) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then, f is said to be *Lipschitz continuous* if there exists some constant $L > 0$ such that f satisfies the inequality

$$\|f(x_1) - f(x_2)\| \leq L \|x_1 - x_2\|, \qquad \forall (x_1, x_2) \in \mathbb{R} \times \mathbb{R}.$$

Prove that the function f is Lipschitz continuous if and only if the total derivative of f is bounded by L, i.e., $\|Df(x)\| \leq L$ for all $x \in \mathbb{R}$.
(Hint 1: This is an "if and only if" statement; thus, you need to prove both directions of the statement.)
(Hint 2: Use the definition of the Total Derivative.)
(Hint 3: When proving that when the Total Derivative is Bounded implies that f is Lipschitz continuous, use the mean-value theorem, i.e., for two points $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$, there exists some point $c \in \mathbb{R}^n$ on the line segment between x_1 and x_2 such that $f(x_1) - f(x_2) = Df(c)^\top (x_1 - x_2)$.)
(Hint 4: When proving that when the Total Derivative is Bounded implies that f is Lipschitz continuous, you will also need to utilize the Cauchy-Schwarz inequality.)