CS57300: Assignment 1

Due date: Sunday September 5, 11:59pm (submit pdf to Brightspace)

Your assignment must be typed and submitted as a PDF. Use of Latex is recommended, but not required. Please complete this assignment independently.

1 Counting (5 pts)

(a) Two cards are drawn from a deck of 52 cards without replacement:

  (i) What is the probability that the second card is a heart, given that the first card is a heart?

  (ii) What is the probability that none of the cards are hearts, given that at most one card is a heart?

(b) One card is selected from a deck of 52 cards and placed in a second deck containing 52 cards. A card is then selected from the second deck.

  (i) What is the probability that a card drawn from the second deck is an ace?

  (ii) If the first card is placed into a deck of 54 cards containing two jokers, then what is the probability that a card drawn from the second deck is an ace?

  (iii) Given that an ace was drawn from the second deck in (ii), what is the conditional probability that an ace was transferred from the first deck?

2 Probability and conditional probability (4 pts)

(a) Suppose that 30 percent of computer owners use an Apple machine, 50 percent use a Windows machine, and 20 percent use Linux. Suppose that 40 percent of Apple users have succumbed to a computer virus, 76 percent of Windows users get the virus, and 55 percent of Linux users get the virus. We select a person at random and learn that their system was infected with the virus. What is the probability that the person is a Windows user?

(b) There are three cards. The first is green on both sides, the second is red on both sides, and the third is green on one side and red on the other. Consider the scenario where a card is chosen at random and one side is shown (also chosen at random). If the side shown is green, what is the probability that the other side is also green?

3 Probability distributions (5 pts)

(a) Let X be a random variable with discrete pdf $f(x) = \frac{x}{8}$ if $x = 1, 2,$ or 5 and zero otherwise.

  (i) Sketch the graph of the discrete pdf $f(x)$.

  (ii) Find $E[X]$ and $Var(X)$.

  (iii) Find $E[2X + 3]$.

(b) The form of the Bernoulli(p) distribution is not symmetric between the two values of X. In some situations, it will be more convenient to use an equivalent formulation for which $x \in \{-1, 1\}$, in which case the distribution can be written as:

$$P(x|p) = \left(\frac{1-p}{2}\right)^{(1-x)/2} \left(\frac{1+p}{2}\right)^{(1+x)/2}$$

Show that this distribution is normalized (i.e., sums to 1) and evaluate its mean and variance.

4 Independence (5 pts)

(a) Prove the following:
If A and B are independent events, then $P(A|B) = P(A)$.

(b) Prove the following:
If A and B are conditionally independent given Z, that is, $P(A, B|Z) = P(A|Z)P(B|Z)$, then $P(A|B, Z) = P(A|Z)$.

(c) A box contains the following four slips of paper, each having exactly the same dimensions: (1) `win prize 1`, (2) `win prize 2`, (3) `win prize 3`, (4) `win prizes 1, 2, and 3`. One slip will be randomly selected. Let A_1 = `win prize 1`, A_2 = `win prize 2`, and A_3 = `win prize 3`. Show that A_1, A_2, and A_3 are *pairwise* independent, but that the three events are not *mutually* independent (i.e., $P(A_1 \wedge A_2 \wedge A_3) \neq P(A_1)P(A_2)P(A_3)$).

5 Expectation (5 pts)

(a) Let $X_1, ..., X_n \sim$ Bernoulli(p=0.5). Let $Y_n = \max\{X_1, ..., X_n\}$.

  (i) Find $E[Y_n]$.

  (ii) Plot $E[Y_n]$ as a function of n.

  (iii) How is the distribution of the max (Y_n) different from that of a single Bernoulli (X_i)?

(b) You and your friend are playing the following game: two dice are rolled; if the total showing is divisible by 4, you pay your friend \$12. If you want to make the game fair, how much should she pay you when the total is not divisible by 4? A fair game is one in which your expected winnings are \$0.

6 Conditional Expectation (4 pts)

Consider the setting where you first roll a fair 6-sided die, and then you flip a fair coin the number of times shown by the die. Let D refer to the outcome of the die roll (i.e., number of coin flips) and let H refer to the number of heads observed after D coin flips.

(a) Suppose the outcome of rolling the fair 6-sided die is d. Determine $E[H|d]$ and $Var(H|d)$ for all possible values of d.

(b) Determine $E[H]$ and $Var(H)$.

7 Covariance and Correlation (6 pts)

(a) Show that if $E[X|Y = y] = c$ for some constant c, then X and Y are uncorrelated.

(b) Show $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$.

(c) Let X_1 and X_2 be quantitative and verbal scores on one aptitude exam and Y_1 and Y_2 be corresponding scores on another exam. If $Cov(X_1, Y_1) = 5$, $Cov(X_1, Y_2) = 1$, $Cov(X_2, Y_1) = 2$, and $Cov(X_2, Y_2) = 8$, what is the covariance between the two total scores $X_1 + X_2$ and $Y_1 + Y_2$?

8 Distance and Correlation Measures (5 pts)

(a) Show how Euclidean distance can be expressed as a function of **cosine similarity** when each data vector has an L_2 length of 1.

(b) Show how Euclidean distance can be expressed as a function of **correlation** when each data point has been standardized by subtracting its mean and dividing by its standard deviation.

9 Linear Algebra (5 pts)

(a) Specify whether the following matrix has an inverse without trying to compute the inverse: (*Recall that a matrix A is invertible if and only if* $\det(A) \neq 0$.)

$$\begin{bmatrix} 9 & 1 & 9 & 9 & 9 \\ 9 & 0 & 9 & 9 & 2 \\ 4 & 0 & 0 & 5 & 0 \\ 9 & 0 & 3 & 9 & 0 \\ 6 & 0 & 0 & 7 & 0 \end{bmatrix}$$

(b) Find eigenvalues and eigenvectors of the following matrix:

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

10 Statistical Inference (6 pts)

Suppose you obtain N data points $X = \{x_1, x_2, ..., x_N\}$ from a normal distribution whose variance is δ^2 and mean is unknown.

(a) What is the maximum likelihood estimation of the normal distribution's mean value μ?

(b) If the prior distribution for μ is a normal distribution with mean value of η and variance of λ^2, i.e., $\mu \sim N(\eta, \lambda^2)$, what is the maximum a-posteriori estimation of μ?