Module 1: PROBABILITY AND STATISTICS - 50 points
Open Book Exam

(Part - A)

Answer the following questions in True/False and provide an appropriate reasoning. [$3 \times 5 =$ 15 points]

1. If two events (both with probability greater than 0) are mutually exclusive, then the two events can be independent.
2. The probability of obtaining heads on the 101st coin toss for a fair coin (P[Heads]=0.5) that resulted in 90 heads in the first 100 coin tosses is 0.9.
3. A low value of R^2 indicates that a large proportion of the variation in y (predicted value) remains unexplained by the regression.

(Part - B)

Answer the following multiple choice questions. Each question has a single correct answer. Provide solution steps. [$3 \times 5 = 15$ points]

1. The table below provides the data from an exam conducted in a class of size 180. Find the probability that a randomly selected girl student will have the highest grade (A).

	Got A grade	Got $< A$ grade
Girl	25	40
Boy	55	60

   (a) 13/36
   (b) 4/9
   (c) 5/36
   (d) 5/13.

2. There two six faced dice named x and y, each with nomenclature $\{1, 2, 3, 4, 5, 6\}$ and $\{1, 1, 3, 3, 5, 5\}$, respectively. Find the mutual information of the outcomes of the two rolls. Use the formula for mutual information:

$$I(x; y) = H(x) - H(x|y)$$

   (a) 0.2
   (b) 0.5
   (c) 0.08
   (d) None of the above.

3. A fair coin is tossed n times and lands heads k times. What is the maximum likelihood estimate for θ the probability of heads.

(a) 0.5

(b) n/k

(c) k/n

(d) none of the above

(Part - C)

Descriptive-answer questions. [$2 \times 10 = 20$ points]

1. Ram wants to buy a used Maruthi car. His colleague is selling his used 2008 model Maruthi for Rs 70000. The market rate for this brand of cars is in fact Rs 80000, if it is in a good shape. Ram does not know much about cars. A mechanic says that he can test it and ascertain its condition. For this he charges Rs 1000. If the car is in bad shape, let us assume it takes Rs.15000 to get it repaired. Ram thinks that the probability of his colleagues car being in good shape is 70%. Calculate the expected net gain from buying his colleagues car without getting it tested by the mechanic. The mechanic and the tests promised by him are not very reliable.

   - P [car passes test—it is in good shape] = 0.8
   - P [car passes test—it is in bad shape] = 0.35

   Should Ram get the car tested by the mechanic before making the decision?

2. We perform a t-test for the null hypothesis $H_0 := 10$ at significance level $\alpha = 0.05$ by means of a dataset consisting of $n = 16$ elements with sample mean 11 and sample variance 4. Should we reject the null hypothesis in favor of $H_A : \mu \neq 10$?

Module 1: LINEAR ALGEBRA AND OPTIMIZATION - 50 points
Open Book Exam

(Part - A)

Answer the following questions in True/False and provide an appropriate reasoning. (15 Points)

(a) If two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then the matrix $\mathbf{A} = \mathbf{u}\mathbf{v}^T + \mathbf{v}\mathbf{u}^T$ is always symmetric.

(b) Given a non-zero matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and a non-zero vector $\mathbf{x} \in \mathbb{R}^n$, then the quantity $c = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}$ is always greater than or equal to 0 i.e $c \geq 0$.

(c) The singular values associated with SVD of a given matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ are always greater than or equal to zero

(d) A function $f(x)$ has x^* as a strict local minimizer, then both $f'(x^*) = 0$ and $f''(x^*) > 0$ always get satisfied.

(e) In stochastic gradient descent approach, gradients corresponding to all data points are computed for every update of parameters.

(Part - B)

Answer the following multiple choice questions. Each question has a single correct answer.

(a) Let $\mathbf{A} = \begin{bmatrix} 2 & 4 & -3 \\ 4 & 2 & -3 \\ -3 & -3 & 9 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 0 & -2 \\ 1 & 0 & 3 \\ 2 & -1 & 1 \end{bmatrix}$. What is the value of tensor quantitative product of $\mathbf{A}$ and $\mathbf{B}$ i.e., $c = A \bullet B = \sum_{i,j} A_{ij}B_{ij}$

   (a) 12

   (b) 10

   (c) 13

   (d) 11

(b) A symmetric matrix $\mathbf{M}$ is defined to be positive definite if $\mathbf{x}^T\mathbf{M}\mathbf{x} > 0$ for all non-zero vectors $\mathbf{x}$ in $\mathbb{R}^n$. Then which of the following is true for a positive definite matrix.

(a) Few eigenvalues of matrix $\mathbf{M}$ can be complex.

(b) The matrix $\mathbf{M}$ can be always decomposed as $\mathbf{M} = \mathbf{R}^T\mathbf{R}$ for an upper triangular matrix $\mathbf{R}$.

(c) Few eigenvalues of the matrix $\mathbf{M}$ can be negative.

(d) None of the eigenvalues of the matrix $\mathbf{M}$ can be negative but few of them can be zero.

(c) In a data science problem, let $f(w, b)$ denote a loss function to be minimized where w and b are the parameters to be found in the learning model. In this model, w is a parameter corresponding to an infrequent feature but an important feature while b is a parameter corresponding to a frequent feature. To this end, which of the following describes the best update rule for an adaptive learning rate method in a gradient descent approach:

(a) Learning rates should be smaller to have larger updates for w.

(b) Learning rates should be bigger to have smaller updates for b.

(c) Learning rates should be smaller for w to have smaller updates for w and larger for b to have larger updates for b.

(d) Learning rates should be higher for w to have larger updates for w and smaller for b to have smaller updates for b.

(Part - C)

Descriptive-answer questions.

(a) Consider the function $f(w, b) = w\exp(-w^2 - b^2)$. The following exercise allows you to set up an optimization algorithm to minimize $f(w, b)$ starting with an initial guess $(w^0, b^0) = (1, 0)$.

(a) For the $k = 0$ iteration, evaluate the gradient descent direction by computing the negative of the gradient of the function $f(w, b)$ at the initial guess $(w^0, b^0) = (1, 0)$. Subsequently, choosing a fixed learning rate $\eta = 0.1$ compute the updates to the parameters (w, b) corresponding to 1^{st} iteration $(k = 1)$ and 2^{nd} iteration $(k = 2)$ of the gradient descent algorithm i.e., compute (w^1, b^1) from (w^0, b^0) and similarly compute (w^2, b^2) from (w^1, b^1).

(b) Compute the update to the parameters (w, b) corresponding to 1^{st} iteration $(k = 1)$ of the Newton's method i.e compute (w^1, b^1) from (w^0, b^0) using Newton's update.

(b) Consider a data having two features and four measurements for each of the features. A data matrix $\mathbf{X} \in \mathbb{R}^{4\times 2}$ is formed with each of these feature vectors forming the two columns of the matrix $\mathbf{X}$ and upon computing the SVD of $\mathbf{X}$, one finds that the matrix $\mathbf{U}$ formed by the left singular vectors to be $\mathbf{U} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}$ and the matrix $\mathbf{V}$ formed by the right singular vectors to be $\mathbf{V} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$. Further the singular values are found to be $\sqrt{2}$ and 0.5

(a) Reconstruct the data matrix $\mathbf{X}$ from the singular values and singular vectors given above. Further compute the Frobenius norm of the data matrix $\mathbf{X}$. What will be the rank of this matrix $\mathbf{X}$? Justify.

(b) Construct the rank one approximation of $\mathbf{X}$ using the dominant singular vectors and the corresponding singular value.

(c) Compute the matrix $\mathbf{C} = \mathbf{X}^T\mathbf{X}$. Verify that the given non-zero singular values of $\mathbf{X}$ are square roots of the eigenvalues of the matrix $\mathbf{C}$ by computing the eigenvalues of the matrix $\mathbf{C}$.

MORE SAMPLE QUESTIONS:

PROBABILITY AND STATISTICS

(Part - A)

Answer the following questions in True/False and provide an appropriate reasoning.

(a) A p-value is the probability that the null hypothesis is true.

(b) The mean of a normalized data is zero.

(c) If two events A and B are independent, then $p(A|B) = p(B|A)$.

(Part - B)

Answer the following multiple choice questions. Each question has a single correct answer.

(a) A medical treatment has a success rate of 0.8. Two patients will be treated with this treatment. Assuming the results are independent for the two patients, what is the probability that neither one of them will be successfully cured?

  (a) 0.5

  (b) 0.36

  (c) 0.2

  (d) 0.04

(b) Suppose that the probability of event A is 0.2 and the probability of event B is 0.4. Also, suppose that the two events are independent. Then $P(A|B)$ is:

  (a) 0.2

  (b) 0.5

  (c) 0.08

  (d) None of the above.

(c) A hypothesis test is done in which the alternative hypothesis is that more than 10% of a population is left-handed. The p-value for the test is calculated to be 0.25. Which statement is correct?

  (a) We can conclude that more than 10% of the population is left-handed.

  (b) We can conclude that more than 25% of the population is left-handed.

  (c) We can conclude that exactly 25% of the population is left handed.

  (d) We cannot conclude that more than 10% of the population is left-handed.

(Part - C)

Descriptive-answer questions.

(a) Alvin throws darts at a circular target of radius r and is equally likely to hit any point in the target. Let X be the distance of Alvin's hit from the center.
(a) Find the PDF, the mean. and the variance of X.
(b) The target has an inner circle of radius t. If $X \leq t$, Alvin gets a score of $S = 1/X$. Otherwise his score is $S = 0$. Find the CDF of S. Is S a continuous random variable?

(b) A person X moves to a new house and she is "fifty-percent sure" that the phone number is 2537267. To verify this, she uses the house phone to dial 2537267. She obtains a busy signal and concludes that this is indeed the correct number. Assuming that the probability of a typical seven-digit phone number being busy at any given time is 1%. what is the probability that X's conclusion was correct?

LINEAR ALGEBRA AND OPTIMIZATION

(Part - A)

Answer the following questions in True/False and provide an appropriate reasoning.

(a) Three column vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are always linearly independent if these three vectors are mutually orthogonal to each other.

(b) The L_2 norm of a vector can sometimes be negative.

(c) The left singular vectors associated with SVD of a given matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ spans the column space of $\mathbf{A}$.

(d) Every local minimum of a function is also a global minimum.

(e) Mini batch gradient descent approach is a second order method

(Part - B)

Answer the following multiple choice questions. Each question has a single correct answer.

(a) Let $\mathbf{A} = \begin{bmatrix} 2 & 4 & -3 \\ 4 & 2 & -3 \\ -3 & -3 & 9 \end{bmatrix}$ and two of its eigenvalues are $12, -2$. What can be a possible third eigenvalue?

(a) 2

(b) 9

(c) 3

(d) 4

(b) Consider the function $f(x) = (x^2 - 1)^3$, the point x = 0 is

(a) a strict local maximum

(b) a saddle point

(c) a strict local minimum

(d) not a stationary point

(c) In a data science problem, let $f(w, b)$ denote a loss function to be minimized where w and b are the parameters to be found in the learning model. In this model, w is a parameter corresponding to an infrequent feature but an important feature while b is a parameter corresponding to a frequent feature. To this end, which of the following describes the best update rule for an adaptive learning rate method in a gradient descent approach:

(a) Learning rates should be smaller to have larger updates for w.

(b) Learning rates should be bigger to have smaller updates for b.

(c) Learning rates should be smaller for w to have smaller updates for w and larger for b to have larger updates for b.

(d) Learning rates should be higher for w to have larger updates for w and smaller for b to have smaller updates for b.

(Part - C)

Descriptive-answer questions.

(a) Consider the function $f(x_1, x_2) = x_1 \exp(-x_1^2 - x_2^2)$. Compute the following.

(a) Evaluate the gradient of the above function f and thereby compute the stationary points by setting this gradient to zero.

(b) Compute the Hessian of the above function f at all the stationary points and verify whether each of these points is a strict local maxima/minima.

(b) Consider the three vectors $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$; $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$; $\mathbf{a}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

(a) Verify the vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are linearly independent.

(b) Construct a matrix $\mathbf{A}$ having $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ as the first, second and third columns of $\mathbf{A}$ respectively. Subsequently compute the QR factorization of this matrix $\mathbf{A}$ using the procedure described in the class.