Quiz 1

Name: **Roll Number**................

- Answer precisely in the space given. No overwriting. Make assumptions, if really ambiguous.
- Questions 1-10: 1 Points; Questions 11-12: 2.5 Points. Total 15 Points. ∘

---

1. Consider a simple linear classifier, with bias. $\mathbf{w} = [2.5, 0.5, 1.5]^T$. This is tested on the following 5 samples. $(\mathbf{x}_i, y_i)$

$$([0,0]^T, +), ([+1,+1]^T, +), ([-1,+1]^T, +), ([-1,-1]^T, -), ([+1,-1]^T, -)$$

What is the accuracy of this classifier ? Ans: ———-%.

2. A 3NN classifier is built over four training examples, and tested on the same test set of Q1.

$$([+2.0, 0.5]^T, +), (-3.0, +0.5]^T, -), ([0.0, 3.0]^T, +), ([0.0, -3.0]^T, -)$$

What is the accuracy of this classifier ? Ans: ———-%.

3. Posterior probability of two classes (i.e., $P(\omega_1|\mathbf{x})$ and $P(\omega_2|\mathbf{x})$) for two classes in a binary classification problem (with $\mathbf{x} \in R^1$)are known to be normal distributions $\mathcal{N}(20, 25)$ and $\mathcal{N}(30, 5)$ respectively. Given a test sample of 25.1, predict the class. Ans: ————-

4. Given the following confusion matrix.

	Predicted +ve	Predicted -ve
Actual +ve	A	B
Actual -ve	C	D

Write an expression for precision. Ans ————

Write an expression for accuracy. Ans ————

5. An $N \times N$ matrix A is composed of consecutive N^2 integers starting from K. (i.e., $K, K+1, \ldots, K+(N^2-1)$)

Rank of A is independent of N. (True or False?) ———

Rank of A is independent of K. (True or False?) ———

6. Bag I contain 10 white and 5 black balls. Bag II contains 15 white and 5 black balls.

A ball is drawn at random from one of the bags, and it is found to be white. What is the probability that it was drawn from Bag I.

Ans ————-

7. A set of samples were pre-processed by a simple linear transformation $\mathbf{x}' = \mathbf{A}\mathbf{x}$ Let d_{ij} is the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ and d'_{ij} is the distance between $\mathbf{x}'_i$ and $\mathbf{x}'_j$

   (a) When $\mathbf{A}$ is a permutation matrix (i.e., every row and column has only one '1' and all other elements being '0'; Note: $\mathbf{A}$ need not be identity). Then, $d_{ij} = d'_{ij}$ for all i, j. True or False? ———-

   (b) When $\mathbf{A}$ is $\rho\mathbf{I}$, a simple linear classifier $sign(\mathbf{w}^T\mathbf{x})$ will not report any change in accuracy after the transformation. True or False? ———

8. Consider a vocabulary of size d. One hot representation of a word i, $\mathbf{w}_i$, is "1" at the location (index) corresponding to that word and zero else where. Given a document that contains P words, $\mathbf{w_1}, \ldots, \mathbf{w_P}$, we compute

$$\mathbf{x} = \sum_{i=1}^{P} \mathbf{w}_i$$

Then,

(a) $\mathbf{x}$ is the histogram of the words, with its ith element x_i as the frequency of i th word.

(b) $\mathbf{x}$ is in R^d independent of the number of words in the document.

(c) $\mathbf{x}$ is in R^P independent of the vocabulary size.

(d) $\sum_i x_i$ is P (x_i is the i th element of $\mathbf{x}$)

Which of the above statements are True? ______________

9. Consider the covariance matrix Σ

(a) Σ is symmetric

(b) Σ is PSD

(c) Σ is Diagonal if the distribution is Normal.

(d) Σ can not be Diagonal if the distribution is Normal.

(e) None of the above are true

Which of the above statements are True? ______________

10. Consider the following statements:

(a) Product of Eigen values is Determinant of a matrix

(b) A matrix of $m \times n$ can have max (m,n) non-zero eigen values

(c) If determinant of a matrix is zero, means at least one of the eigen value is zero.

(d) Eigen vectors are orthogonal to each other (i.e., $\mathbf{v_1}^T\mathbf{v_2} = 0$)

(e) All the above are true

Which of the above statements are True? ______________

11. We saw the loss function for linear regression as $J(\theta) = (Y - X\theta)^T(Y - X\theta)$ We saw that we get a closed form solution for θ by solving $\frac{\partial J(\theta)}{\partial \theta} = 0$:

$$\frac{\partial}{\partial \theta}\left(Y^TY - 2Y^TX\theta + \theta^TX^TX\theta\right) = 0; \implies -2X^TY + 2X^TX\theta = 0; \implies \theta = (X^TX)^{-1}X^TY$$

Now find the closed form solution that minimizes this loss function (assume A is symmetric):

$$J(\theta) = (Y - X\theta)^TA(Y - X\theta)$$

(a) $\theta = (X^TAX)^{-1}X^TY$
(b) $\theta = (X^TX)^{-1}X^TAY$
(c) $\theta = (X^TAX)^{-1}X^TAY$
(d) $\theta = (X^TAX)^{-1}X^TA^{-1}Y$
(e) None of these

Write all correct options ——————-

12. Consider the function

$$f(w) = w^2 + w + 1$$

We want to find the minima of the function using gradient descent. We start at $w^0 = 5.0$. Write update equation for computing w^{k+1} from w^k. Ans: —————-

What should be the learning rate η so that we reach the minima in a single step?

Ans: ——————

Rough Work (will not be graded)

Quiz 1

Name: **Roll Number**................

- Answer precisely in the space given. No overwriting. Make assumptions, if really ambiguous.
- Questions 1-10: 1 Points; Questions 11-12: 2.5 Points. Total 15 Points. ‡

1. Consider a simple linear classifier, with bias. $\mathbf{w} = [2.5, 0.5, 1.5]^T$. This is tested on the following 5 samples. $(\mathbf{x}_i, y_i)$

$$([0,0]^T, +), ([+1,+1]^T, +), ([-1,+1]^T, +), ([-1,-1]^T, -), ([+1,-1]^T, -)$$

How many errors this classifier makes? Ans: ———-.

2. A 3NN classifier is built over four training examples, and tested on the same test set of Q1.

$$([+2.0, 0.5]^T, +), (-3.0, +0.5]^T, -), ([0.0, 3.0]^T, +), ([0.0, -3.0]^T, -)$$

What is the accuracy of this classifier ? Ans: ———-%.

3. Posterior probability of two classes (i.e., $P(\omega_1|\mathbf{x})$ and $P(\omega_2|\mathbf{x})$) for two classes in a binary classification problem (with $\mathbf{x} \in R^1$)are known to be normal distributions $\mathcal{N}(30, 25)$ and $\mathcal{N}(40, 5)$ respectively. Given a test sample of 35.1, predict the class. Ans: ————-

4. Given the following confusion matrix.

	Predicted +ve	Predicted -ve
Actual +ve	A	B
Actual -ve	C	D

Write an expression for recall. Ans ——————

Write an expression for accuracy. Ans ——————

5. An $N \times N$ matrix A is composed of consecutive N^2 integers starting from K. (i.e., $K, K+1, \ldots, K + (N^2 - 1)$)

Rank of A is independent of N. (True or False?) ———

Rank of A is independent of K. (True or False?) ———

6. Bag I contain 5 white and 10 black balls. Bag II contains 5 white and 15 black balls.

A ball is drawn at random from one of the bags, and it is found to be white. What is the probability that it was drawn from Bag I.

Ans ————-

7. A set of samples were pre-processed by a simple linear transformation $\mathbf{x}' = \mathbf{A}\mathbf{x}$ Let d_{ij} is the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ and d'_{ij} is the distance between $\mathbf{x}'_i$ and $\mathbf{x}'_j$

   (a) When $\mathbf{A}$ is a permutation matrix (i.e., every row and column has only one '1' and all other elements being '0'; Note: $\mathbf{A}$ need not be identity). Then, $d_{ij} = d'_{ij}$ for all i, j. True or False? ———-

(b) When $\mathbf{A}$ is $\rho\mathbf{I}$, a simple linear classifier $sign(\mathbf{w}^T\mathbf{x})$ will not report any change in accuracy after the transformation. True or False? ———

8. Consider a document is represented by a histogram of the words in the document. $\mathbf{h}$ i.e., h_i is the number of occurrence of the i th word in the document.

   We define a linguistic operation: Paraphrasing (P1). P1 is defined as permuting sentences in a document and rewriting a sentence by permuting the words.

   (a) $\mathbf{h}$ is invariant to the P1

   (b) $\mathbf{h}$ is not invariant to the P1

   (c) $\mathbf{h}$ is invariant under in which order the vocabulary is constructed (eg. "a to z" or "z to a")

   (d) a Euclidean distance computed over $\mathbf{h}_i$ and $\mathbf{h}_j$ is invariant under in which order the vocabulary is constructed (eg. "a to z" or "z to a".)

   Which of the above statements are True? ——————

9. Consider the covariance matrix Σ

   (a) Σ is symmetric

   (b) Σ is PSD

   (c) Σ is Diagonal if the distribution is Normal.

   (d) Σ can not be Diagonal if the distribution is Normal.

   (e) None of the above are true

   Which of the above statements are True? ——————

10. Consider the following statements:

    (a) Product of Eigen values is Determinant of a matrix

    (b) A matrix of $m \times n$ can have max (m,n) non-zero eigen values

    (c) If determinant of a matrix is zero, means at least one of the eigen value is zero.

    (d) Eigen vectors are orthogonal to each other (i.e., $\mathbf{v_1}^T\mathbf{v_2} = 0$)

    (e) All the above are true

    Which of the above statements are True? ——————

11. We saw the loss function for linear regression as $J(\theta) = (Y - X\theta)^T(Y - X\theta)$ We saw that we get a closed form solution for θ by solving $\frac{\partial J(\theta)}{\partial \theta} = 0$:

$$\frac{\partial}{\partial \theta}\left(Y^TY - 2Y^TX\theta + \theta^TX^TX\theta\right) = 0; \implies -2X^TY + 2X^TX\theta = 0; \implies \theta = (X^TX)^{-1}X^TY$$

Now find the closed form solution that minimizes this loss function (assume A is symmetric):

$$J(\theta) = (Y - X\theta)^TA(Y - X\theta)$$

(a) $\theta = (X^TAX)^{-1}X^TY$

(b) $\theta = (X^TX)^{-1}X^TAY$

(c) $\theta = (X^TAX)^{-1}X^TAY$

(d) $\theta = (X^TAX)^{-1}X^TA^{-1}Y$

(e) None of these

Write all correct options ————————-

12. Consider the function

$$f(w) = w^2 + w + 1$$

We want to find the minima of the function using gradient descent. We start at $w^0 = -5.0$. Write update equation for computing w^{k+1} from w^k. Ans: —————-

What should be the learning rate η so that we reach the minima in a single step?

Ans: ——————

Rough Work (will not be graded)

Quiz 1

Name: **Roll Number**................

- Answer precisely in the space given. No overwriting. Make assumptions, if really ambiguous.
- Questions 1-10: 1 Points; Questions 11-12: 2.5 Points. Total 15 Points. †

---

1. Consider a simple linear classifier, with bias. $\mathbf{w} = [2.5, 0.5, 1.5]^T$. This is tested on the following 5 samples. $(\mathbf{x}_i, y_i)$

$$([0,0]^T, +), ([+1,+1]^T, +), ([-1,+1]^T, +), ([-1,-1]^T, -), ([+1,-1]^T, -)$$

What is the accuracy of this classifier ? Ans: ———-%.

2. A 3NN classifier is built over four training examples, and tested on the same test set of Q1.

$$([+2.0, 0.5]^T, +), (-3.0, +0.5]^T, -), ([0.0, 3.0]^T, +), ([0.0, -3.0]^T, -)$$

How many errors this classifier makes ? Ans: ———-.

3. Posterior probability of two classes (i.e., $P(\omega_1|\mathbf{x})$ and $P(\omega_2|\mathbf{x})$) for two classes in a binary classification problem (with $\mathbf{x} \in R^1$)are known to be normal distributions $\mathcal{N}(20, 5)$ and $\mathcal{N}(30, 25)$ respectively. Given a test sample of 24.9, predict the class. Ans: ————-

4. Given the following confusion matrix.

	Predicted +ve	Predicted -ve
Actual +ve	A	B
Actual -ve	C	D

Write an expression for accuracy. Ans ————

Write an expression for precision. Ans ————

5. An $N \times N$ matrix A is composed of consecutive N^2 integers starting from K. (i.e., $K, K+1, \ldots, K+(N^2-1)$)

Rank of A is independent of N. (True or False?) ———

Rank of A is independent of K. (True or False?) ———

6. Bag I contain 15 white and 5 black balls. Bag II contains 10 white and 5 black balls.

A ball is drawn at random from one of the bags, and it is found to be white. What is the probability that it was drawn from Bag I.

Ans ————-

7. A set of samples were pre-processed by a simple linear transformation $\mathbf{x}' = \mathbf{A}\mathbf{x}$ Let d_{ij} is the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ and d'_{ij} is the distance between $\mathbf{x}'_i$ and $\mathbf{x}'_j$

   (a) When $\mathbf{A}$ is a permutation matrix (i.e., every row and column has only one '1' and all other elements being '0'; Note: $\mathbf{A}$ need not be identity). Then, $d_{ij} = d'_{ij}$ for all i, j. True or False? ———-

   (b) When $\mathbf{A}$ is $\rho\mathbf{I}$, a simple linear classifier $sign(\mathbf{w}^T\mathbf{x})$ will not report any change in accuracy after the transformation. True or False? ———

8. Consider a document is represented by a histogram of the words in the document $\mathbf{h}$ i.e., h_i is the number of occurrence of the i th word in the document.

   We define a linguistic operation: Paraphrasing (P2). P2 is defined as replacing a set of words by their respective synonym.

   (a) $\mathbf{h}$ is invariant to the P2
   (b) $\mathbf{h}$ is not invariant to the P2
   (c) $\mathbf{h}$ is invariant under in which order the vocabulary is constructed (eg. "a to z" or "z to a")
   (d) a Euclidean distance computed over $\mathbf{h}_i$ and $\mathbf{h}_j$ is invariant under in which order the vocabulary is constructed (eg. "a to z" or "z to a")

   Which of the above statements are True? ____________

9. Consider the covariance matrix Σ

   (a) Σ is symmetric
   (b) Σ is PSD
   (c) Σ is Diagonal if the distribution is Normal.
   (d) Σ can not be Diagonal if the distribution is Normal.
   (e) None of the above are true

   Which of the above statements are True? ____________

10. Consider the following statements:

   (a) Product of Eigen values is Determinant of a matrix
   (b) A matrix of $m \times n$ can have max (m,n) non-zero eigen values
   (c) If determinant of a matrix is zero, means at least one of the eigen value is zero.
   (d) Eigen vectors are orthogonal to each other (i.e., $\mathbf{v_1}^T\mathbf{v_2} = 0$)
   (e) All the above are true

   Which of the above statements are True? ____________

11. We saw the loss function for linear regression as $J(\theta) = (Y - X\theta)^T (Y - X\theta)$ We saw that we get a closed form solution for θ by solving $\frac{\partial J(\theta)}{\partial \theta} = 0$:

$$\frac{\partial}{\partial \theta}\left(Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta\right) = 0; \implies -2X^T Y + 2X^T X\theta = 0; \implies \theta = (X^T X)^{-1} X^T Y$$

Now find the closed form solution that minimizes this loss function (assume A is symmetric):

$$J(\theta) = (Y - X\theta)^T A (Y - X\theta)$$

(a) $\theta = (X^T A X)^{-1} X^T Y$
(b) $\theta = (X^T X)^{-1} X^T A Y$
(c) $\theta = (X^T A X)^{-1} X^T A Y$
(d) $\theta = (X^T A X)^{-1} X^T A^{-1} Y$
(e) None of these

Write all correct options ——————-

12. Consider the function

$$f(w) = w^2 + w + 1$$

We want to find the minima of the function using gradient descent. We start at $w^0 = 5.0$.
Write update equation for computing w^{k+1} from w^k. Ans: ————-
What should be w^1 if the learning rate $\eta = 0.1$?
Ans: ——————

Rough Work(will not be graded)

Quiz 1

Name: **Roll Number**................

- Answer precisely in the space given. No overwriting. Make assumptions, if really ambiguous.
- Questions 1-10: 1 Points; Questions 11-12: 2.5 Points. Total 15 Points.

---

1. Consider a simple linear classifier, with bias. $\mathbf{w} = [2.5, 0.5, 1.5]^T$. This is tested on the following 5 samples. $(\mathbf{x}_i, y_i)$

   $$([0,0]^T, +), ([+1,+1]^T, +), ([-1,+1]^T, +), ([-1,-1]^T, -), ([+1,-1]^T, -)$$

   How many errors this classifier makes ? Ans: ———-.

2. A 3NN classifier is built over four training examples, and tested on the same test set of Q1.

   $$([+2.0, 0.5]^T, +), (-3.0, +0.5]^T, -), ([0.0, 3.0]^T, +), ([0.0, -3.0]^T, -)$$

   How many errors this classifier makes? Ans: ———-.

3. Posterior probability of two classes (i.e., $P(\omega_1|\mathbf{x})$ and $P(\omega_2|\mathbf{x})$) for two classes in a binary classification problem (with $\mathbf{x} \in R^1$)are known to be normal distributions $\mathcal{N}(30, 5)$ and $\mathcal{N}(40, 25)$ respectively. Given a test sample of 34.9, predict the class. Ans: ————-

4. Given the following confusion matrix.

	Predicted +ve	Predicted -ve
Actual +ve	A	B
Actual -ve	C	D

   Write an expression for accuracy. Ans ——————

   Write an expression for recall. Ans ——————

5. An $N \times N$ matrix A is composed of consecutive N^2 integers starting from K. (i.e., $K, K + 1, \ldots, K + (N^2 - 1)$)

   Rank of A is independent of N. (True or False?) ———

   Rank of A is independent of K. (True or False?) ———

6. Bag I contain 5 white and 15 black balls. Bag II contains 5 white and 10 black balls.

   A ball is drawn at random from one of the bags, and it is found to be white. What is the probability that it was drawn from Bag I. Ans ————-

7. A set of samples were pre-processed by a simple linear transformation $\mathbf{x}' = \mathbf{A}\mathbf{x}$ Let d_{ij} is the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ and d'_{ij} is the distance between $\mathbf{x}'_i$ and $\mathbf{x}'_j$

   (a) When $\mathbf{A}$ is a permutation matrix (i.e., every row and column has only one '1' and all other elements being '0'; Note: $\mathbf{A}$ need not be identity). Then, $d_{ij} = d'_{ij}$ for all i, j. True or False? ———-

   (b) When $\mathbf{A}$ is $\rho\mathbf{I}$, a simple linear classifier $sign(\mathbf{w}^T\mathbf{x})$ will not report any change in accuracy after the transformation. True or False? ———

8. Consider a vocabulary of size d. One hot representation of a word i, $\mathbf{w}_i$, is "1" at the location (index) corresponding to that word and zero else where. Given a document that contains P words, $\mathbf{w_1}, \ldots, \mathbf{w_P}$, we compute

$$\mathbf{x} = \sum_{i=1}^{P} \mathbf{w}_i$$

Then,

(a) $\mathbf{x}$ is the histogram of the words, with its ith element x_i as the frequency of i th word.

(b) $\mathbf{x}$ is in R^d independent of the number of words in the document.

(c) $\mathbf{x}$ is in R^P independent of the vocabulary size.

(d) $\sum_i x_i$ is P (x_i is the i th element of $\mathbf{x}$)

Which of the above statements are True? ____________

9. Consider the covariance matrix Σ

(a) Σ is symmetric

(b) Σ is PSD

(c) Σ is Diagonal if the distribution is Normal.

(d) Σ can not be Diagonal if the distribution is Normal.

(e) None of the above are true

Which of the above statements are True? ____________

10. Consider the following statements:

(a) Product of Eigen values is Determinant of a matrix

(b) A matrix of $m \times n$ can have max (m,n) non-zero eigen values

(c) If determinant of a matrix is zero, means at least one of the eigen value is zero.

(d) Eigen vectors are orthogonal to each other (i.e., $\mathbf{v_1}^T \mathbf{v_2} = 0$)

(e) All the above are true

Which of the above statements are True? ____________

11. We saw the loss function for linear regression as $J(\theta) = (Y - X\theta)^T (Y - X\theta)$ We saw that we get a closed form solution for θ by solving $\frac{\partial J(\theta)}{\partial \theta} = 0$:

$$\frac{\partial}{\partial \theta}\left(Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta\right) = 0; \implies -2X^T Y + 2X^T X\theta = 0; \implies \theta = (X^T X)^{-1} X^T Y$$

Now find the closed form solution that minimizes this loss function (assume A is symmetric):

$$J(\theta) = (Y - X\theta)^T A (Y - X\theta)$$

(a) $\theta = (X^T A X)^{-1} X^T Y$
(b) $\theta = (X^T X)^{-1} X^T A Y$
(c) $\theta = (X^T A X)^{-1} X^T A Y$
(d) $\theta = (X^T A X)^{-1} X^T A^{-1} Y$
(e) None of these

Write all correct options ————————-

12. Consider the function

$$f(w) = w^2 + w + 1$$

We want to find the minima of the function using gradient descent. We start at $w^0 = -5.0$. Write update equation for computing w^{k+1} from w^k. Ans: —————-

What should be w^1 if the learning rate $\eta = 0.1$?

Ans: ————————

Rough Work(will not be graded)