Course › Final E... › Final E... › Audit Fi...

Audit Final Exam

Q1

0/3 points (graded)

A particular data set has 4 possible labels, with the following frequencies:

Label	Frequency
A	50%
B	20%
C	20%
D	10%

a) What is the error rate of a classifier that picks a label (A,B,C,D) at random, each with probability $1/4$? Give your answer as a number in the range $[0, 1]$.

0 ✖ **Answer:** 0.75

0

b) One very simple type of classifier just returns the same label, always. What label should it return?

0 ✖ **Answer:** A

c) What is the error rate of the classifier from b)? Give your answer as a number in the range $[0, 1]$.

0 ✖ **Answer:** 0.5

0

Submit You have used 5 of 5 attempts

ℹ Answers are displayed within the problem

Q2

0/1 point (graded)

We decide to use 4-fold cross-validation to figure out the right value of k to choose when running k-nearest neighbor on a data set of size 10,000. When checking a particular value of k, we look at four different training sets. What is the size of each of these training sets?

2500 ✖ **Answer:** 7500

2500

Submit You have used 5 of 5 attempts

ℹ Answers are displayed within the problem

Q3

0/3 points (graded)

For the point $x = (1, 2, 3, 4)$ in $\mathbb{R}^4$, compute the following.

a) $\|x\|_1$

✖ **Answer:** 10

b) $\|x\|_2$

✖ **Answer:** 5.477

c) $\|x\|_\infty$

0

✗ **Answer:** 4

0

Submit You have used 5 of 5 attempts

ℹ Answers are displayed within the problem

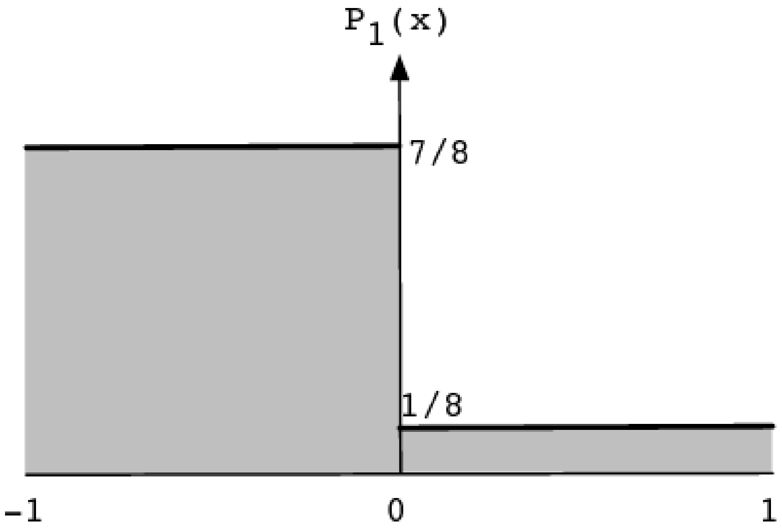
Q4

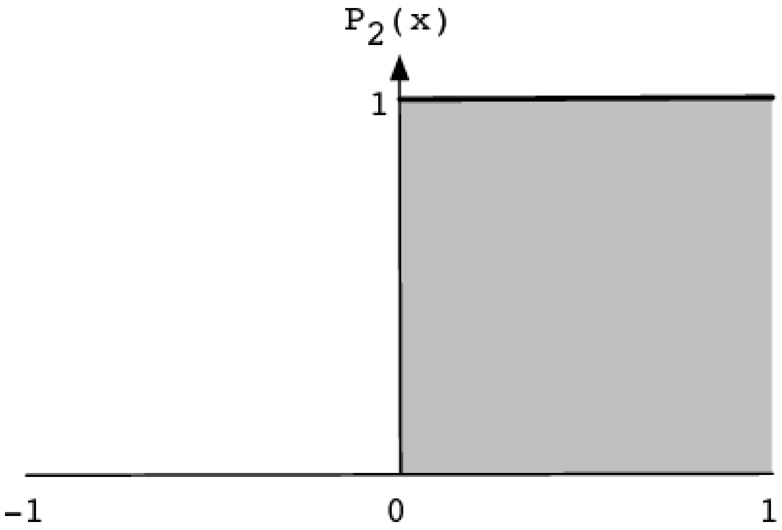
0/2 points (graded)

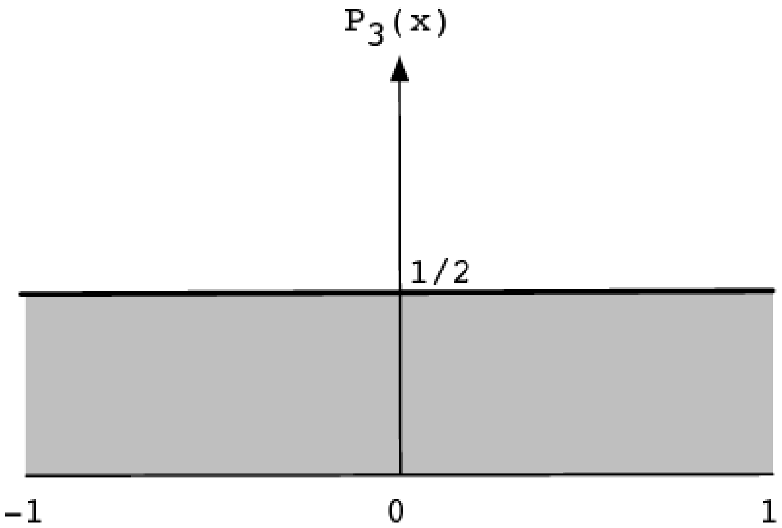
Suppose we have one-dimensional data points lying in $X = [-1, 1]$, that have associated labels in $Y = \{1, 2, 3\}$. The individual classes have weights

$$\pi_1 = \frac{1}{3}, \quad \pi_2 = \frac{1}{6}, \quad \pi_3 = \frac{1}{2}$$

and densities P_1, P_2, P_3 as shown below. (For instance, P_1 is the density of the points whose label is 1; in particular, this means that P_1 integrates to 1.)







Based on this information, what labels should be assigned to the following points?

a) $-1/2$

✖ **Answer:** 1

b) $1/2$

0

✖ **Answer:** 3

0

Explanation

In each case (for each of the two given values of x), we need to compute $\pi_1 P_1(x), \pi_2 P_2(x), \pi_3 P_3(x)$, and then select the label that has the largest value.

Submit

You have used 5 of 5 attempts

ℹ Answers are displayed within the problem

Q5

0/2 points (graded)

Random variables X, Y take on values in the range $\{-1, 0, 1\}$ and have the following joint distribution.

		Y		
		-1	0	1
	-1	0	0	$1/3$
X	0	0	$1/3$	0
	1	$1/3$	0	0

a) What is the covariance between X and Y?

✖ **Answer:** -2/3

b) What is the correlation between X and Y?

0

✖ **Answer:** -1

0

You have used 5 of 5 attempts

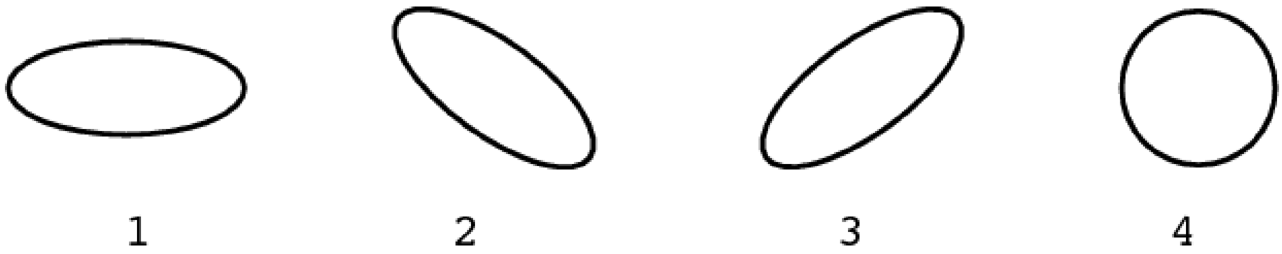
Submit

Answers are displayed within the problem

Q6

3.0/3.0 points (graded)

Here are four possible shapes of Gaussian distributions:



For each of the following Gaussians $N(\mu, \Sigma)$, indicate which of these shapes (1,2,3,4) is the best approximation.

a) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$

1 ✓ **Answer:** 1

1

b) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 9 & 2 \\ 2 & 1 \end{pmatrix}$

3 ✓ **Answer:** 3

3

c) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & -0.75 \\ -0.75 & 1 \end{pmatrix}$

2 ✓ **Answer:** 2

2

Submit You have used 3 of 5 attempts

Answers are displayed within the problem

Q7

1/3 points (graded)

We have n data points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$ and we store them in a matrix X, one point per row.

a) True or false: X has dimension $d \times n$.

○ True

● False

✓

b) True or false: $X^T X$ has dimension $d \times d$.

○ True ✓

● False

✗

c) Which of the following is a matrix with (i, j) entry $x^{(i)} \cdot x^{(j)}$?

○ XX

○ $X^T X$

○ XX^T ✓

● $X^T X^T$

✗

Explanation

For part (c), notice that the first and last options aren't even valid products, unless $d = n$. Of the middle two, $X^T X$ has dimension $d \times d$ while XX^T has dimension $n \times n$.

Submit You have used 2 of 2 attempts

ℹ Answers are displayed within the problem

Q8

2.0/3.0 points (graded)

Suppose we solve a classification problem with k classes by using a Gaussian generative model in which the jth class is specified by parameters π_j, μ_j, Σ_j. In each of the following situations, say whether the decision boundary is `linear`, `spherical`, or `other quadratic`.

a) We compute the empirical covariance matrices of each of the k classes, and then set $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k$ to the **average** of these matrices.

linear ▾ ✓ **Answer:** linear

b) The covariance matrices Σ_j are all **diagonal**, but no two of them are the same.

spherical ▾ ✗ **Answer:** other quadratic

c) There are two classes (that is, $k = 2$) and the covariance matrices Σ_1 and Σ_2 are multiples of the identity matrix.

spherical ▾ ✓ **Answer:** spherical

Submit You have used 2 of 2 attempts

ℹ Answers are displayed within the problem

Q9

2/4 points (graded)

Suppose that we have data points $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$, where $x^{(i)}, y^{(i)} \in \mathbb{R}$, and that we want to fit them with a line that passes through the origin. The general form of such a line is $y = ax$: that is, the sole parameter is $a \in \mathbb{R}$.

a) In this setting, what are the **predictor** and **response** variables?

predictor x and response y ▾ ✓ **Answer:** predictor x and response y

b) The goal is to find the value of a that minimizes the squared error on the data. We will do this by first writing down a **loss function** $L(\cdot)$. Which of the following statements is an accurate description of the loss function? Select all that apply.

☐ It takes a parameter a and returns a real number. ✓

☐ It takes a data set and returns a parameter a.

☐ It is based on the given data set. ✓

☑ It is the same regardless of the data set.

✗

c) Using calculus, find the optimal setting of a. The answer is of the form $a = N/D$ where the numerator N and the denominator D can be found in the following list.

$\sum_{i=1}^{n} \left(y^{(i)} - x^{(i)}\right) x^{(i)}$

$\sum_{i=1}^{n} x^{(i)} y^{(i)}$

$\sum_{i=1}^{n} {y^{(i)}}^2$

$\sum_{i=1}^{n} {x^{(i)}}^2$

$\sum_{i=1}^{n} \left(y^{(i)} - x^{(i)}\right)^2$

Which of these is N and which is D?

$N =$

○ $\sum_{i=1}^{n} \left(y^{(i)} - x^{(i)}\right) x^{(i)}$

○ $\sum_{i=1}^{n} x^{(i)} y^{(i)}$ ✓

○ $\sum_{i=1}^{n} {y^{(i)}}^2$

● $\sum_{i=1}^{n} {x^{(i)}}^2$

○ $\sum_{i=1}^{n} \left(y^{(i)} - x^{(i)}\right)^2$

✗

$D =$

○ $\sum_{i=1}^{n} \left(y^{(i)} - x^{(i)}\right) x^{(i)}$

○ $\sum_{i=1}^{n} x^{(i)} y^{(i)}$

○ $\sum_{i=1}^{n} {y^{(i)}}^2$

● $\sum_{i=1}^{n} {x^{(i)}}^2$

○ $\sum_{i=1}^{n} \left(y^{(i)} - x^{(i)}\right)^2$

✓

Explanation

For a line $y = ax$, the total squared loss on the n data points is:

$L(a) = \sum_{i=1}^{n} \left(y^{(i)} - ax^{(i)}\right)^2.$

To minimize this, we take the derivative with respect to a:

$\frac{dL}{da} = -2 \sum_{i} \left(y^{(i)} - ax^{(i)}\right) x^{(i)}$

and then set this to zero, to get:

$a = \frac{\sum_i y^{(i)} x^{(i)}}{\sum_i {x^{(i)}}^2}.$

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Q10

1/1 point (graded)

When learning a logistic regression model from training data $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$, which of the following do we try to do? Select all that apply.

☐ Maximize the probabilities of the $x^{(i)}$

☑ Maximize the conditional probabilities of the $y^{(i)}$ given $x^{(i)}$

☐ Maximize the joint probabilities of $x^{(i)}$ and $y^{(i)}$

✓

Submit You have used 2 of 2 attempts

✓ Correct (1/1 point)

Q11

0/2 points (graded)

Given a set of data points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, we want to find the vector $w \in \mathbb{R}^d$ that minimizes this loss function:

$$L(w) = \sum_{i=1}^{n} \left(w \cdot x^{(i)}\right) + \frac{1}{2} c \|w\|^2.$$

Here $c > 0$ is some constant.

a) Let s denote the sum of the data points, that is, $s = \sum_{i=1}^{n} x^{(i)}$. Express $\nabla L(w)$ in terms of s, c, and w.

○ $\nabla L(w) = s + w$

○ $\nabla L(w) = s + cw$✓

● $\nabla L(w) = cw$

○ $\nabla L(w) = s/c + w$

✗

b) What value of w minimizes $L(w)$? Give the answer in terms of s and c.

○ $w = -\frac{s}{c}$ ✓

● $w = cs$

○ $w = \frac{s}{4c}$

○ $w = -\frac{s}{2c}$

✗

Submit You have used 2 of 2 attempts

ℹ Answers are displayed within the problem

Q12

0/2 points (graded)

For some fixed vector $u \in \mathbb{R}^d$, define

$$F(x) = \|x - u\|^2.$$

We wish to determine whether $F(x)$ is a convex function of x.

a) The Hessian matrix $H(x)$ is of the form cI, where I is the $d \times d$ identity matrix and c is some constant. What is c?

0 ✗ **Answer:** 2

0

b) Is $F(x)$ a convex function?

○ Yes ✓

● No

○ It depends on the specific vector u

✗

Explanation

For the first part, we have
$F(x) = \sum_{j=1}^{d}(x_j - u_j)^2$.
Thus
$\frac{dF}{dx_j} = 2(x_j - u_j)$
and $d^2F/dx_k dx_j$ is either 2 if $j = k$ or 0 otherwise. Thus the Hessian is $2I$, which is PSD, implying that F is convex.

Submit You have used 2 of 2 attempts

ℹ Answers are displayed within the problem

Q13

1/1 point (graded)

A particular line in $\mathbb{R}^2$ passes through the points $(0, 1)$ and $(2, 0)$ and is specified by equation $w \cdot x + b = 0$, where $b = -2$ and $w \in \mathbb{R}^2$. What is w?

○ $w = (0, 1)$

○ $w = (0, 2)$

● $w = (1, 2)$

○ $w = (2, -1)$

✓

Submit You have used 1 of 2 attempts

✓ Correct (1/1 point)

Q14

1/1 point (graded)

The dual form of the hard-margin SVM returns a vector α. Which data points $x^{(i)}$ are the support vectors in this solution?

○ Those with $\alpha_i = 0$

● Those with $\alpha_i > 0$

○ Those with $\alpha_i \geq 0$

○ The support vectors cannot be determined simply by looking at α

✓

Submit You have used 1 of 2 attempts

✓ Correct (1/1 point)

Q15

1/1 point (graded)

Which of the following are quadratic functions of $x = (x_1, x_2, x_3)$? Select all that apply.

☐ $x_1^2 + x_2^2 + x_3^2 + x_1 x_2 x_3$

☑ $x_1 x_2 + x_1 x_3 + x_2 x_3$

☐ $(x_1 + x_2 + x_1 x_2)^2$

☑ $1 + x_1 + x_2 + x_3 + 10x_2^2$

✓

Submit You have used 2 of 2 attempts

✓ Correct (1/1 point)

Q16

0.0/1.0 point (graded)

In order to solve the kernel SVM optimization problem, what information about the data set $\{(x^{(i)}, y^{(i)})\}$ do we need to provide to the optimization procedure?

● The labels $y^{(i)}$ and the basis expansions $\Phi\left(x^{(i)}\right)$

○ The labels $y^{(i)}$ and the squared norms $\Phi\left(x^{(i)}\right) \cdot \Phi\left(x^{(i)}\right)$

○ The labels $y^{(i)}$ and the pairwise dot products $\Phi\left(x^{(i)}\right) \cdot \Phi\left(x^{(j)}\right)$ ✓

✗

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Q17

1/4 points (graded)

Consider the following data set consisting of five points in $\mathbb{R}^1$:

$$-10, -8, 0, 8, 10.$$

We would like to cluster these points into $k = 3$ groups. Determine the optimal k-means solution.

a) What is the location of the leftmost center?

-8 ✗ **Answer:** -9

-8

b) What is the location of the middle center?

0 ✓ **Answer:** 0

0

c) What is the location of the rightmost center?

8 ✗ **Answer:** 9

8

d) What is the k-means cost of this optimal solution?

✗ **Answer:** 4

Submit You have used 4 of 4 attempts

Answers are displayed within the problem

Q18

0/1 point (graded)

Which of the following are reasons for which hierarchical clustering might be preferred to flat clustering? Select all that apply.

- ☐ It does not require the number of clusters to be specified. ✓
- ☑ It captures the structure of the data at multiple scales. ✓
- ☑ It is computationally simpler.
- ☐ It has an intuitive cost function for which an optimal solution can efficiently be obtained.

✗

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Q19

0/1 point (graded)

What is the projection of the vector $(3, 5, -9)$ onto the direction $(0.6, -0.8, 0)$?

0 ✖ **Answer:** -2.2

0

Submit You have used 2 of 2 attempts

🛈 Answers are displayed within the problem

Q20

0/2 points (graded)

For a particular four-dimensional data set, the top two eigenvectors of the covariance matrix are:

$$\frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \ \frac{1}{2}\begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}.$$

a) What is the PCA projection of point $(2, 4, 2, 6)$ into two dimensions? Write it in the form (a, b).

○ $(2, 2)$

○ $(2, 3)$

○ $(7, 3)$ ✓

● $(4, 6)$

✖

b) What is the reconstruction, from this projection, to a point in the original four-dimensional space? Write it in the form (a, b, c, d)

○ $(2,5,2,5)$ ✓

○ $(2,1,2,2)$

● $(4,2,2,2)$

○ $(2,6,2,4)$

✗

Submit You have used 2 of 2 attempts

ℹ Answers are displayed within the problem

Q21

0/2 points (graded)

Consider the \(2 \times 2\) matrix \(M = \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}\).

a) One of its eigenvectors is \(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}\). What is the corresponding eigenvalue?

0 ✗ **Answer:** 6

\(\)

b) Its other eigenvector is \(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}\). What is the corresponding eigenvalue?

0 ✗ **Answer:** 4

\(\)

Submit You have used 5 of 5 attempts

ℹ Answers are displayed within the problem

Q22

0/1 point (graded)

A feedforward neural network has five layers, each consisting of \(100\) nodes, and each fully connected to the previous layer. Roughly how many parameters does this network have?

○ \(100\)

○ \(500\)

● \(10000\)

○ \(50000\) ✓

✗

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Q23

0/1 point (graded)

It is known that any function over \(d\) variables can be arbitrarily well approximated by:

○ A linear function

○ A neural net with one hidden layer containing \(d\) nodes

○ A neural net with one hidden layer containing potentially a large number of nodes ✓

● A neural net with depth \(d\), in which each hidden layer has \(d\) nodes

✗

Submit You have used 2 of 2 attempts

Answers are displayed within the problem