

## Quiz 2

● Graded

Student

Brian Bertness

Total Points

29 / 30 pts

Question 1

Q1

11 / 12 pts

1.1

(a)

3 / 3 pts

✓ + 3 pts Everything correct (give full points if the student showed correct answer only)

+ 1 pt Correct Answer

+ 2 pts Correct Formula

+ 0 pts Unattempted/Incorrect

- 1 pt Arithmetic error

1.2

(b)

4 / 5 pts

+ 5 pts Everything correct

✓ + 3 pts Correct formulation for Covariance

✓ + 2 pts Make sure to use unbiased (n-1)

✓ - 1 pt Arithmetic Error

- 1 pt Improper Scaling for Data

+ 0 pts Unattempted/Incorrect

1

$X = x - \mu$

1.3

(c)

4 / 4 pts

✓ + 4 pts Everything correct

+ 2 pts Specify the meaning of entry (covariance/correlation between corresponding feature)

+ 2 pts Specify the meaning of value (e.g. The value is 0 for  $S_{22}$  because the first feature doesn't change)

+ 0 pts Unattempted/Incorrect

## Question 2

Q2

10 / 10 pts

2.1 (a)

5 / 5 pts

✓ + 5 pts Everything correct

+ 2.5 pts Correct vector for PCA ( $\langle -1, 1 \rangle$  vector)

+ 2.5 pts Correct vector for LDA ( $\langle 1, 1 \rangle$  vector)

+ 0 pts Incorrect answers / No attempt to draw the vectors

2.2 (b)

5 / 5 pts

✓ + 5 pts Correct choice and explanation (given the directions from part a). The expected explanation includes at least one of the following: (1) LDA takes into account class labels in finding vector to project to so it is usually better; (2) projecting with PCA will mix the data points from different classes so it is usually not as good

+ 3.5 pts Correct choice with error in explanation. E.g., (1) PCA uses class information in *projection*; it uses the class label to find *the vector to project the data to*; (2) vague points like LDA is more informative of the class label/its easier to separate classes using LDA; (3) LDA creates the largest difference in class means or maximizes separation between classes; (4) LDA gives thresholds for easier classification

+ 2.5 pts Partial credit for the attempt to the question.

+ 0 pts No attempt/completely wrong

## Question 3

Q3

8 / 8 pts

✓ + 8 pts Everything correct

+ 2 pts Select (a)

+ 2 pts Select (b)

+ 2 pts Not select (c)

+ 2 pts NOT select (d)

+ 0 pts Incorrect/unattempted

No questions assigned to the following page.

## CSCI 5521: Machine Learning Fundamentals (Spring 2024)

## Quiz 2 (Thurs, Feb 22)

Due on Gradescope at 02:00 PM, Friday, Feb 23

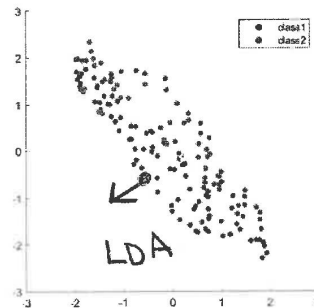
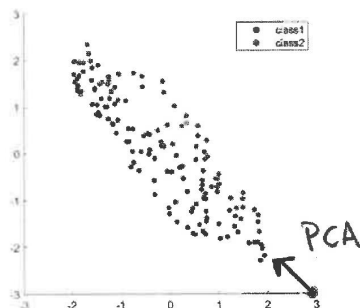
## Instructions:

- This quiz has 3 questions, 30 points, on 2 page.
- Please write your name & ID on this cover page.

1. (12 points) For three data points  $\begin{pmatrix} -3 \\ 0 \end{pmatrix}$ ,  $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ ,  $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$ ,

- Derive the sample mean.
- Derive the **unbiased** sample covariance matrix.
- Explain one of the diagonal entries in the covariance matrix (e.g., if your  $\sigma_{11} = c$ , please intuitively explain why it is equal to c here).

2. (10 points) In the following figures, (a) draw the first principal component direction in the left figure, and the first linear discriminant direction in the right figure. Briefly explain.



(b) We are going to perform a binary classification on the data in the reduced 1-D space. Shall we project the data onto the direction found by PCA or LDA? Briefly explain.

Questions assigned to the following page: [1.1](#) and [1.2](#)

Brian Bertness

1478201

$$1a) \text{ mean} = \begin{bmatrix} (-3 + -1 + -2) / 3 \\ 0 / 3 \end{bmatrix}$$

$$= \begin{bmatrix} -6 / 3 \\ 0 / 3 \end{bmatrix}$$

$$= \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$b) X = \begin{bmatrix} -3 - 2 & -1 - 2 & -2 - 2 \\ 0 - 0 & 0 - 0 & 0 - 0 \end{bmatrix}$$

$$= \begin{bmatrix} -5 & -3 & -4 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\Sigma = \text{cov}(X) = \frac{1}{n-1} X X^T = \frac{1}{3-1} \begin{bmatrix} -5 & -3 & -4 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -5 & 0 \\ -3 & 0 \\ -4 & 0 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 25 + 9 + 16 & 0 + 0 + 0 \\ 0 + 0 + 0 & 0 + 0 + 0 \end{bmatrix}$$

$$= \begin{bmatrix} 25 & 0 \\ 0 & 0 \end{bmatrix}$$

Question assigned to the following page: [1.3](#)

Brian Bertness

1478201

1.c) Consider  $\begin{bmatrix} 25 & 0 \\ 0 & 0 \end{bmatrix}$

$\sigma_{12} + \sigma_{21}$  Both  $\sigma_{12}$  and  $\sigma_{21}$  represent how much the  $x_1$  and  $x_2$  components of the three data points change in regards to each other. That is,  $x_1$  and  $x_2$  do not vary together or increase/decrease in tandem.

$\sigma_{11}$  The value,  $\sigma_{11}$ , is the variance of the  $x_1$  data.

$\sigma_{22}$  The value,  $\sigma_{22}$ , is the variance of the  $x_2$  data.

Since our original data is  $\begin{bmatrix} -3 \\ 0 \end{bmatrix}$ ,  $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$ ,  $\begin{bmatrix} -2 \\ 0 \end{bmatrix}$

it is clear that the  $x_2$  components do not vary as they are all 0, hence  $\sigma_{22} = 0$ . The  $x_1$  components do vary, thus  $\sigma_{11} = 25$ . Lastly, it is clear that  $x_1$  and  $x_2$  do not vary together.



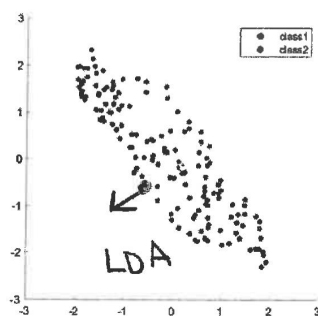
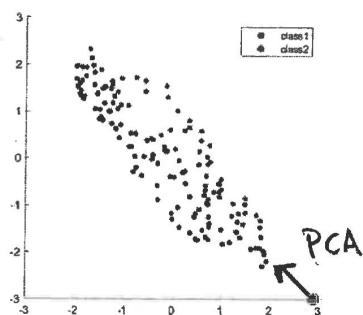
Question assigned to the following page: [2.1](#)

Brian Bertness

1478201

2a)

2. (10 points) In the following figures, (a) draw the first principal component direction in the left figure, and the first linear discriminant direction in the right figure. Briefly explain.



For PCA projection we ignore the class labels. We observe that the direction of maximum variance is the one indicated in the figure.

For LDA we want the means of the classes to be separate and the within class scatter to be a minimum.

Question assigned to the following page: [2.2](#)

Brian Bertness

1478201

2. b) PCA projection ignores the class labels while LDA does not. Therefore, if we want a binary classification according to the labels in the diagram I would choose to project the data in the direction found by LDA.

Question assigned to the following page: [3](#)

Brian Bertness

1478201

3. Both

(a) - k-means + EM always find local optimum

(Note!!! I am assuming the algorithm converges!)

and

(b) - The number of clusters of EM and k-means both need to be manually set by the user.

are correct about k-means and EM for Gaussian Mixtures.