Name+Email ID Brian Bertness 1478201

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**.

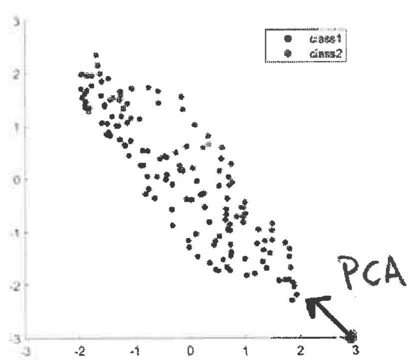
1. (**12 points**) For three data points $\begin{pmatrix} -3 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$,

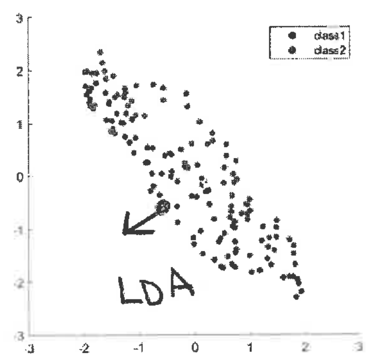
   (a) Derive the sample mean.

   (b) Derive the **unbiased** sample covariance matrix.

   (c) 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.

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

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

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

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

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

Brian Bertness
1478201

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

$\sigma_{12} + \sigma_{21}$ Both σ_{12} and σ_{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.

σ_{11} The value, σ_{11}, is the variance of the X_1 data.

σ_{22} The value, σ_{22}, is the variance of the X_2 data.

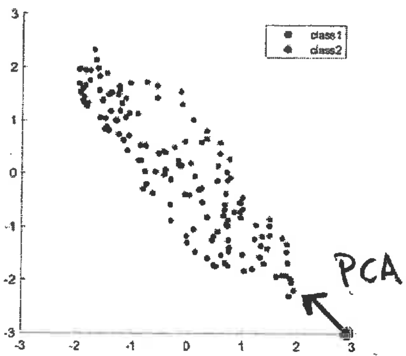
Since our original data is $\begin{bmatrix} -3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \end{bmatrix}$

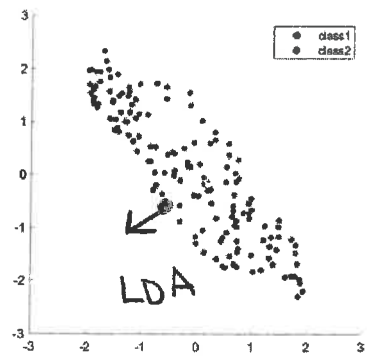
it is clear that the X_2 components do not vary as they are all 0, hence $\sigma_{22} = 0$. The X_1 component do vary, thus $\sigma_{11} = 25$. Lastly, it is clear that X_1 and X_2 do not vary together.

Brian Bertness

2a) 1478201

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.

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.

Brian Bertness
1478201

3. Both
(a) -k-meas + 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.