Name+Email ID

CSCI 5521: Machine Learning Fundamentals (Spring 2024)

Quiz 2 (Thurs, Feb 22)

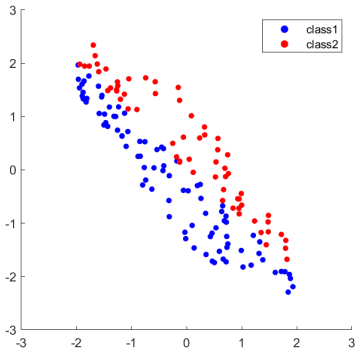
Due on Gradescope at 02:00 PM, Friday, Feb 23

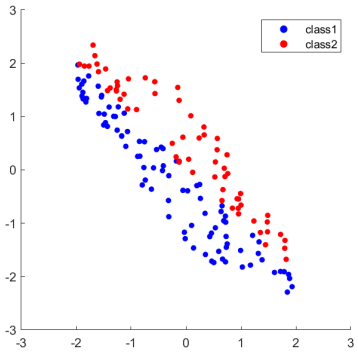
Instructions:

- 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.

3. (**8 points**) Select all the option(s) that are correct about K-means and EM for Gaussian Mixtures:

   (a) K-means and EM always find local optimum.

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

   (c) Both EM and K-means employ hard assignment.

   (d) K-means assumes that the underlying clusters in the data must be distributed according to Gaussian distributions.