Name+Email ID Jon-Michael Hoang | hoang339

CSCI 5521: Machine Learning Fundamentals (Spring 2022)

Quiz 2 (Thurs, Feb 24))

Due on Gradescope at 02:00 PM, Friday, Feb 25

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 $\begin{pmatrix}0\\10\end{pmatrix}, \begin{pmatrix}0\\5\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}$,

   (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 $\sigma_{11} = c$, intuitively explain why it is equal to c here).

unbiased sample cov. mat. = $\begin{bmatrix}0 & 0\\ 0 & 12.5\end{bmatrix}$

a) $\begin{bmatrix}0 & 10\\ 0 & 5\\ 0 & 0\end{bmatrix}$ (x, y) $\Rightarrow x = \frac{0+0+0}{3}$, $y = \frac{10+5+0}{3} \Rightarrow \begin{bmatrix}0\\5\end{bmatrix}$

b) $\begin{matrix} & x & y\\ x & \text{Cov}(x,x) & \text{Cov}(x,y)\\ y & \text{Cov}(y,x) & \text{Var}(y)\end{matrix}$ $= \text{Var}(x) = \frac{0}{3-1} = 0$

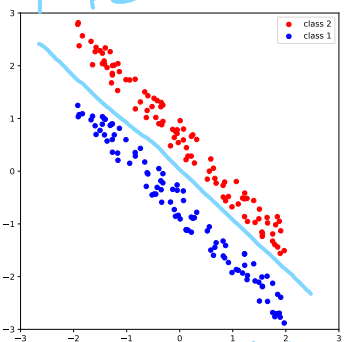
$\text{Var}(y) = \frac{(10-5)^2 + (5-5)^2 + 0^2}{3-1} = 12.5$

$\text{Var}(x,y) = \frac{(0)(5) + (0)(0) + (0)(0)}{3-1} = 0$

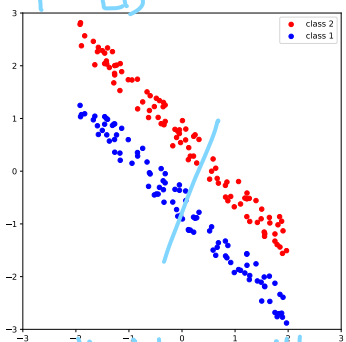
$\text{Var}(y,x) = \frac{(5)(0) + 0 + 0}{3-1} = 0$

c) The reason as to why $\sigma_{00} = 0$ is b/c all points lie about the y-axis - i.e. the line of fit goes directly through the entire y-axis: (0,10), (0,5), (0,0) $\Rightarrow \text{var}(x) = 0 = \sigma_{00}$

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.



1ST PC



1ST LD

Principle Components aims to capture as much of the original variance in the data as possible whereas Linear Discriminants aims to find a low dimensional space s.t. when x is projected, classes are well-separated.

(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.

We use LDA b/c PCA does not consider class labels, and only maximizes the variance in the data. LDA does, however, consider class labels.

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

   (a) They are both clustering methods.

   (b) EM for Gaussian Mixtures has less parameters than K-means.

   (c) EM for Gaussian Mixtures is faster to compute than K-means.

   (d) K-means always finds local optimum while EM for Gaussian Mixtures always finds global optimum.

K-means has faster convergence speed. GMM Mixtures takes more iterations

$r, \mu = 2$

μ, Σ, π

$=$

3

$\Downarrow$

EM has more than K-means.

false - can get saddle