Fall 2018
EAS 595 Project
Due December 15 (11:59 PM)
You can work in a group of 2

In an experiment involving 1000 participants, we recorded two different measurement (F_1 and F_2) while participants performed 5 different tasks ($C_1, C_2, \cdots C_5$). The two measurements are independent and for each class they can be considered to have a normal distribution as follow:

$$P(F_1|C_i) = N(m_{1i}, \sigma_{1i}^2) \text{ and } P(F_2|C_i) = N(m_{2i}, \sigma_{2i}^2) \text{ for } i = 1,2,\cdots 5$$

where, m_{1i}, σ_{1i}^2 are the mean and variance of F_1 for the ith class. Similarly, m_{2i}, σ_{2i}^2 are the mean and variance of F_2 for the ith class.

The goal of this project is to construct a classifier such that for any given values of F_1 and F_2, it can predict the performed task ($C_1, C_2, \cdots C_5$). Let's assume that the classifier calculate the probability of each class given the measurement data, and output the most probable class as the predicted class.

$$Predicted\ Class = \text{argmax}[P(C_i|X)]\ , i = 1,2,\cdots 5$$

The file 'data.m' contains measurements F1 and F2 that are both matrices with the size of 1000x5. Each column contains the information of one of the subjects and each row corresponds to one of the tasks (1st row: 1st task, 2nd row: 2nd task, etc.)

To find the best classifier, perform the following tasks:

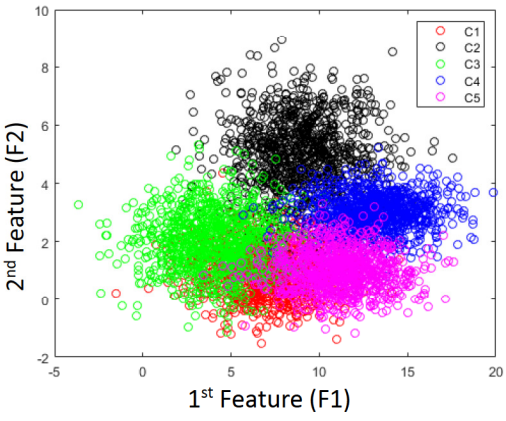
- **Step 1.Training:** Use the data of the first 100 subjects to estimate m_{1i}, σ_{1i}^2 and m_{2i}, σ_{2i}^2

- **Step 2.1.Testing:** Assume that $X = F_1$. Using the Bayes' theorem, calculate the probability of each class for data of the remaining subjects (columns 101-1000 of F_1) and consequently predict the class for each data point. Note that each subject performed 5 different tasks so you need to predict the class of 4500 data points.

- **Step 2.2.Calculating the accuracy of the classifier:** You need to check the percentage of the data whose class are correctly predicted. The true class is the row number of the data. So if you classify F_1(3,131) as class 3, it is correctly classified otherwise you have wrongly predict the class.

  Classification accuracy = correct predictions / total predictions (which is 4500 in this case)
  Error rate = incorrect predictions / total predictions

- **Step 3. Standard Normal (Z-Score):** Assume F_1 to be a subjective measure. In this case the mean value and the range of data reported by one subject will not be consistent with another subject. In other to remove the effect of individual differences, you have to normalize the data of each subject using the standard normal formulation (removing the mean and dividing by standard deviation). Calculate Z_1 (the standard normal of F_1) and plot the distribution of the data using Z_1 and F_2, and compare it to the distribution in F_1 and F_2 shown on right.



- **Step 4.** Repeat 2.1 and 2.2 for the following cases:
    - Case 2: $X = Z_1$ (Note for this case you need to repeat the training step as well)
    - Case 3: $X = F_2$
    - Case 4: $X = \begin{bmatrix} Z_1 \\ F_2 \end{bmatrix}$. Note that this is a multivariate normal distribution and you need to use the independence assumption.

- **Step 5.** Compare the classification rate of the four cases.

- **Step 6.** Write a report in IEEE format (maximum 2 pages) summarizing and discussing the results. Particularly, identify which case works the best and why.

Note:

- You need to submit your code and report on UBlearn by the end of day of 12/15/18
- You can find the IEEE template at the following link https://www.ieee.org/conferences/publishing/templates.html