Name:
ID:
Date:
ITU, Computer Engineering Dept.
BLG454E, Learning From Data HW1
Due. **March 9, 2017, 23:00 through Ninova. NO LATE SUBMISSION WILL BE ACCEPTED. DO-NOT SUBMIT VIA E-MAIL.**

Instructors: Yusuf Yaslan (yyaslan@itu.edu.tr)
Assistant : Resul Tugay(tugayr@itu.edu)

Policy:
You are encouraged to discuss the questions with your class mates, but the code and the homework you submitted must be on behalf of your team. **Cheating is highly discouraged** for it could mean a zero or negative grade from the homework.
If a question is not clear, please let me know (via email or in class). Unless we indicate otherwise, do not use libraries for machine learning methods. When in doubt, email me.

In order to be able to take the final exam for BLG454E you have to have a **weighted average score of 40 (over 100) for midterm, homeworks and term project.** Otherwise you will get a VF from the course.

Q1-0.5pt)A secret government agency has developed a scanner which determines whether a person is a terrorist. The scanner is fairly reliable; 95% of all scanned terrorists are identifed as terrorists, and 95% of all upstanding citizens are identified as such. An informant tells the agency that exactly one passenger of 100 aboard an aeroplane in which you are seated is a terrorist. The agency decide to scan each passenger and the shifty looking man sitting next to you is the first to test positive. What are the chances that this man is a terrorist?

Q2-0.5pt) For a novel input x, a predictive model of the class c is given by p(c=1|x) = 0.7, p(c=2|x) = 0.2, p(c=3|x) = 0.1. The corresponding utility matrix U(c^{true},c^{pred}) has elements:

$$\begin{bmatrix} 5 & 3 & 1 \\ 0 & 4 & -2 \\ -3 & 0 & 10 \end{bmatrix}$$

In terms of maximal expected utility, which is the best decision to take ?
Q3-1pt) Suppose the weights of randomly selected American female college students are normally distributed with unknown mean μ and standard deviation σ. A random

sample of 10 American female college students yielded the following weights (in pounds):

```
115    122    130    127    149    160    152    138 149    180
```

Based on the definitions given above, identify the likelihood function and the maximum likelihood estimator of μ and σ^2, the mean weight and variance of all American female college students. Using the given sample, find a maximum likelihood estimate of μ as well.

Matlab Task (Write codes in Matlab and provide your plot and results in report**)**
Q4-3pt) You will use the following dataset given for this hw. The last column of the file represents the label (class 0 or class 1)
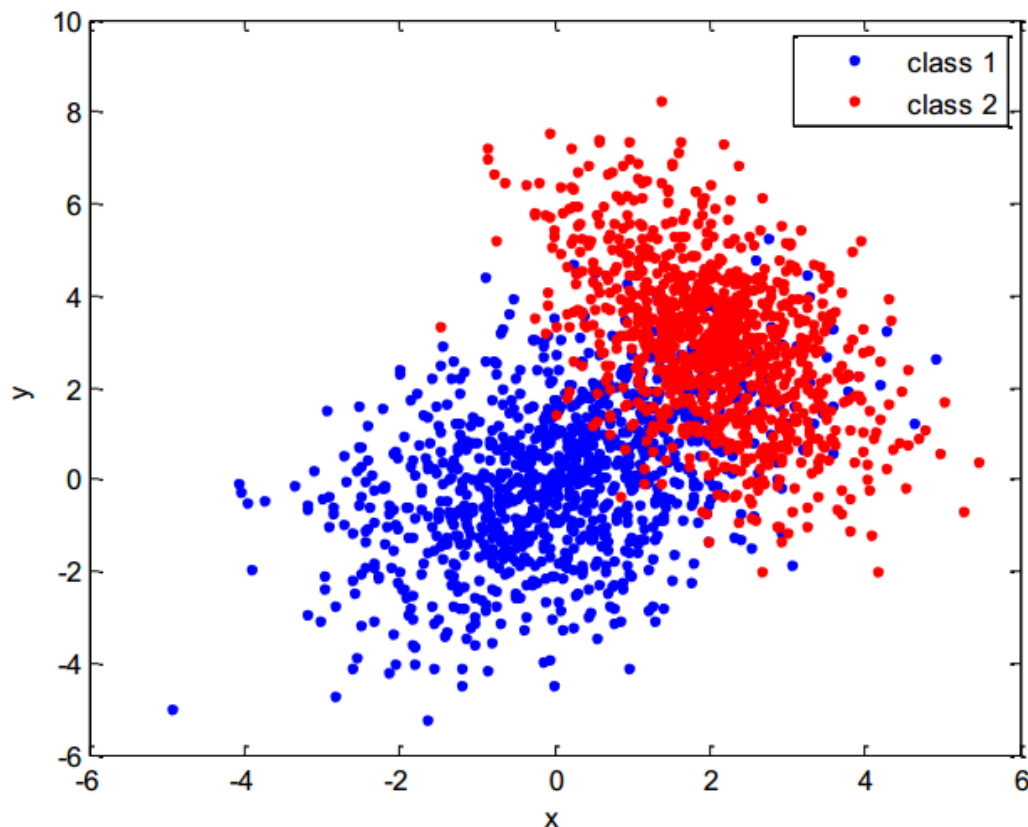
Examine the dataset. The number of features and the number of classes. Classify given dataset using KNN classifier. Use at least two KNN (i.e KNN2 and KNN5) and compare the differences. Write comments below codes you implemented. Plot the classification results and put it into your report.