**Name:**

**ID:**

**Date:**

**ITU, Computer Engineering Dept.**

**BLG527E, Machine Learning HW3**

Due: November 24, 2016, 23:00 through Ninova.

# Instructors:), Yusuf Yaslan (yyaslan@itu.edu.tr)

**Grading:** You must complete the table below according to what you expect to get out of each question.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Q1 | Q2 | Q3 | Total |
| Grade | Max | 1 | 2 | 2 | 5 pts |
| Expected |  |  |  |  |

# Policy:

# Please do your homeworks on your own. You are encouraged to discuss the questions with your class mates, but the code and the hw you submitted must be your own work. Cheating is highly discouraged for it could mean a zero or negative grade from the homework.

# If a question is not clear, please let us know (via email or in class). Unless we indicate otherwise, do not use libraries for machine learning methods. When in doubt, email us.

There will be 5 homeworks this term. Each hw is worth 5 points and each question will be evaluated on a 0/1 basis.

# In order to be able to take the final exam for BLG527E you have to have a weighted average score of 30 (over 100) for midterm and homeworks. Otherwise you will get a VF from the course.

**DO NOT SUBMIT YOUR HOMEWORKS VIA E-MAIL!**

**QUESTIONS**

**In your report you should write your findings neatly.**

**Q1) You will use the Swiss dataset for this hw. The last column of the file shows the label (class -1 or class 1)** [You need to write down the PCA code yourself, do not use a library pca() function. Do not use cov() function, but you may use eig() function]

a) use PCA on class-1, class1 alone and

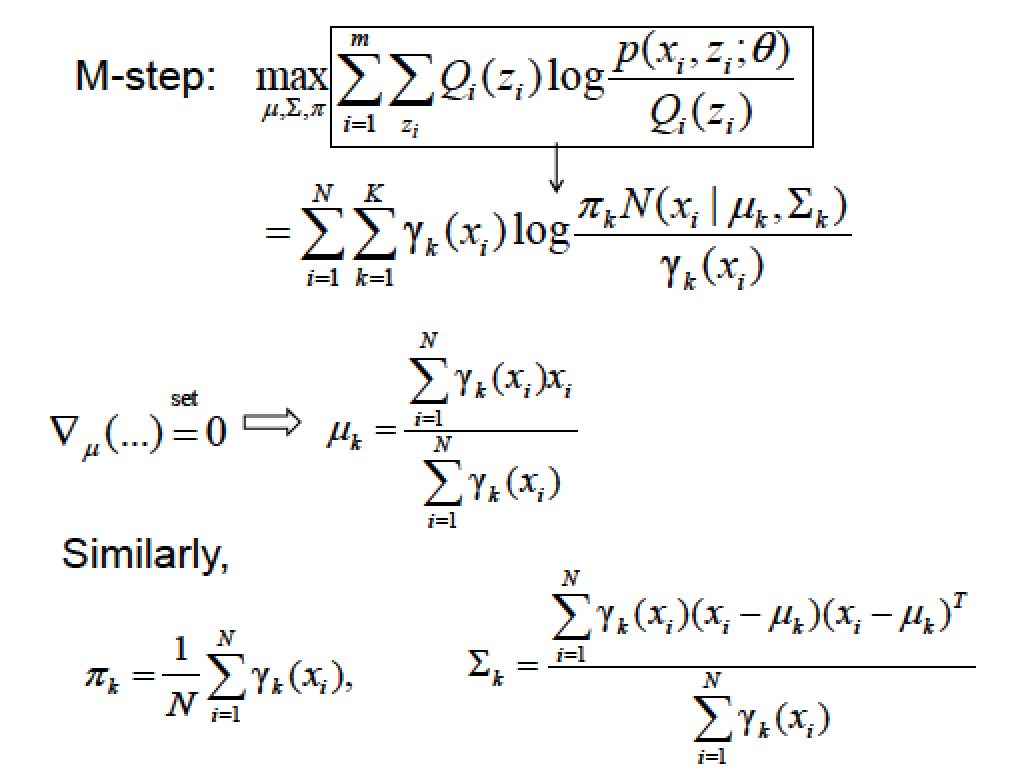
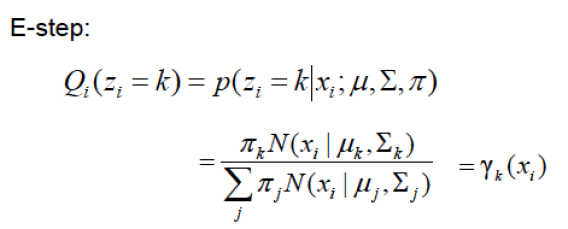
b) use PCA on the whole dataset.

**For both cases, project the instances into two dimensional space and plot the projections in your report.**

Why are these two projections different?

Which one of these two PCA approaches are feasible for a real problem? Answer by considering how you would classify a new instance using the projections you obtained in a) and b) and using KNN5 classifier. The KNN5 (5 nearest neighbor classifier) labels a data point x as follows: Find the nearest (in terms of Euclidean distance) 5 points to x from the training set, label x as the majority of its 5 nearest neighbors’ label. (You should evaluate classification results using 10 fold cross validation)

**Q2)** Write a program that clusters the q3.mat data using the EM on Gaussian Mixture Models with K=2, 3, 4 clusters. **Plot the clustering results.** What is the best number of clusters? Use the following update equations from Bishop and Ng slides (MoG\_EM-Part2.pdf):

****

**Q3)** We are given two coins A and B where the probabilities for heads are *QA* and *QB*. Each coin is selected randomly with a probability p(zk = 1) = πk (k=1,2 and ). After a coin is selected we take 10 observations from that coin. Suppose that you are given the following observations:



1. Write down the expectation maximization steps to find the parameters. **Show the derivation of the formulas that find the parameters (E and M steps).**
2. Write a program that computes the parameters of the given experiment. **Give the parameters in your report**.

**Hint:** Use Binomial distribution.

(See <http://ai.stanford.edu/~chuongdo/papers/em_tutorial.pdf> for detailed descriptions of the sample experiment)