Name:
ID:
Date:
ITU, Computer Engineering Dept.
BLG527E, Machine Learning HW3
Due: November 2, 2014, 22:00 through Ninova.
Instructors: Zehra Cataltepe (cataltepe@itu.edu.tr), 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	Q4	Total
Grade	Max	1	1	1	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.

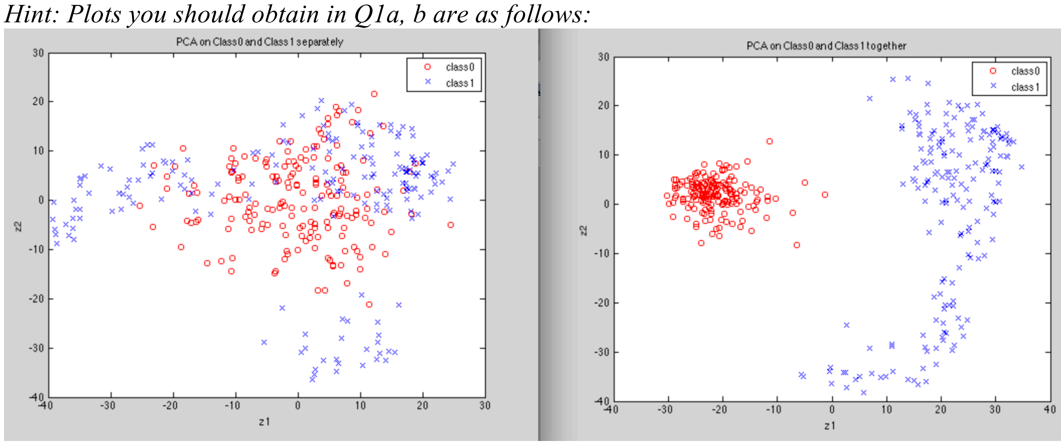
There will be 5 homeworks this term. Each hw is worth 5 points and each question will be evaluated on a 0/1 basis.

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.

QUESTIONS

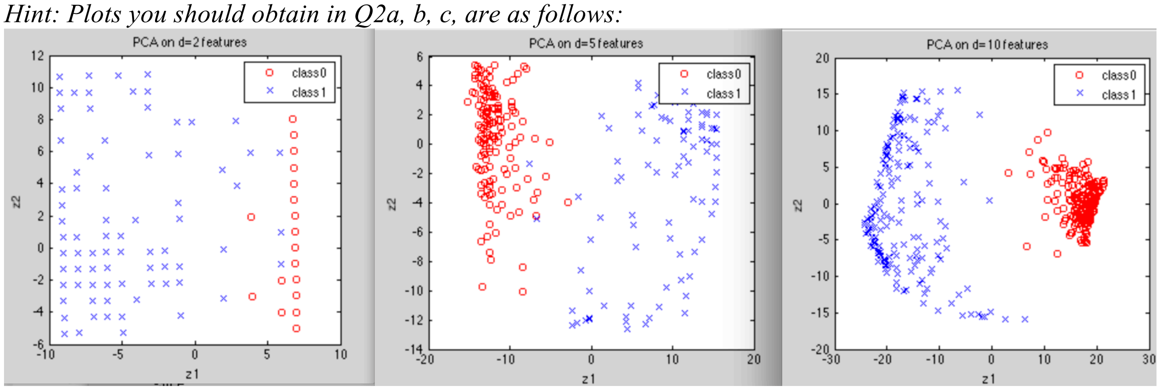
You will use the same subset of (optdigits01) of the optdigits dataset by Alpaydin and Kaynak for this hw. The last column of the file shows the label (class 0 or class 1)

Q1) [You need to write down the PCA code yourself, do not use a library pca() function.]
a) use PCA on class0, class1 alone and
b) use PCA on the whole dataset.
For both cases, project the instances into two dimensional space.
Hint: Plots you should obtain in Q1a, b are as follows:



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.

Q2) Use mrmr algorithm with MID, but modify the code in mrmr_mid_d.m[1] file so that you use abs(corr()) instead of mutualinfo(). Select a) 2, b)5, c)10 features and do feature projection on all data (as you did in Q1b). Plot the projected instances again.
Did feature selection improve the classification accuracy of PCA? Answer using the KNN5 classification method on training data.
Hint: Plots you should obtain in Q2a, b, c, are as follows:



Q3) Cluster the projections of all the data you obtained in Q2c) using the K-means algorithm with K=2, 3, 4 clusters. What is the number of clusters that results in the minimum reconstruction error E?

---

[1] http://www.mathworks.com/matlabcentral/fileexchange/14608-mrmr-feature-selection--using-mutual-information-computation-

Q4) Cluster the projections of all the data you obtained in Q2c) using the EM on Gaussian Mixture Models with K=2, 3, 4 clusters. What is the best number of clusters in terms of the reconstruction error E? Use the following update equations from Bishop and Ng slides (MoG_EM-Part2.pdf):

E-step:

$$Q_i(z_i = k) = p(z_i = k | x_i; \mu, \Sigma, \pi)$$

$$= \frac{\pi_k N(x_i \mid \mu_k, \Sigma_k)}{\sum_j \pi_j N(x_i \mid \mu_j, \Sigma_j)} = \gamma_k(x_i)$$

M-step: $\max_{\mu,\Sigma,\pi} \boxed{\sum_{i=1}^{m} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i; \theta)}{Q_i(z_i)}}$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} \gamma_k(x_i) \log \frac{\pi_k N(x_i \mid \mu_k, \Sigma_k)}{\gamma_k(x_i)}$$

$$\nabla_\mu(\ldots) \overset{\text{set}}{=} 0 \Longrightarrow \mu_k = \frac{\sum_{i=1}^{N} \gamma_k(x_i) x_i}{\sum_{i=1}^{N} \gamma_k(x_i)}$$

Similarly,

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} \gamma_k(x_i), \qquad \Sigma_k = \frac{\sum_{i=1}^{N} \gamma_k(x_i)(x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^{N} \gamma_k(x_i)}$$