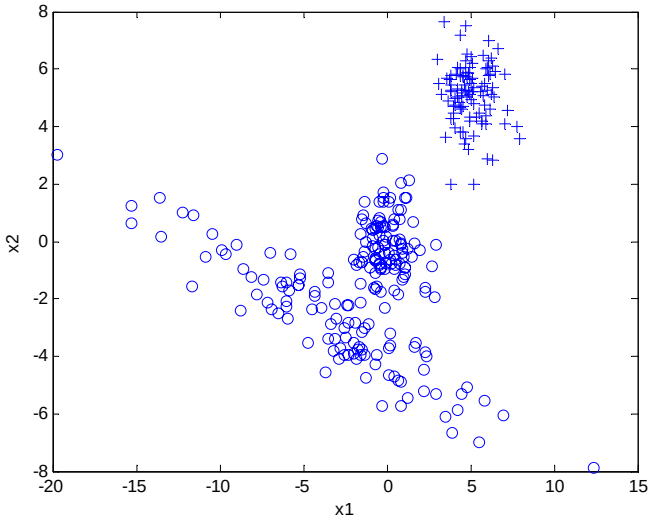
Midterm

Istanbul Technical University- Fall 2006
Pattern Recognition and Analysis (BBL514E)
Machine Learning (BLG527E)

Total worth: 25% of your grade.
Date: Wednesday, November 8, 2006.
Time: 120 mins
Good luck!

1.
1a) [15 pts] What is the VC (Vapnik-Chervonenkis) dimension of a hypothesis class? How and why is it used?
1b) [10 pts] What is k-fold cross-validation? How and why is it used?

2.
2a) [10 pts] Compare and contrast clustering and a dimensionality reduction technique (for example, principal component analysis) in terms of how they transform a certain set of unlabeled data set $X=\{\underline{x}^t\}_{t=1}^N$ where each $\underline{x}^t \in R^d$.
2b) [15 pts] Given the following set of labeled data with two classes (+, o), discuss which input preprocessing techniques you would use. Would you do your input pre-processing on all data or data on a specific class? Assume that your final classifier is a hyperplane, give a sketch of how you would construct your classifier.



3)

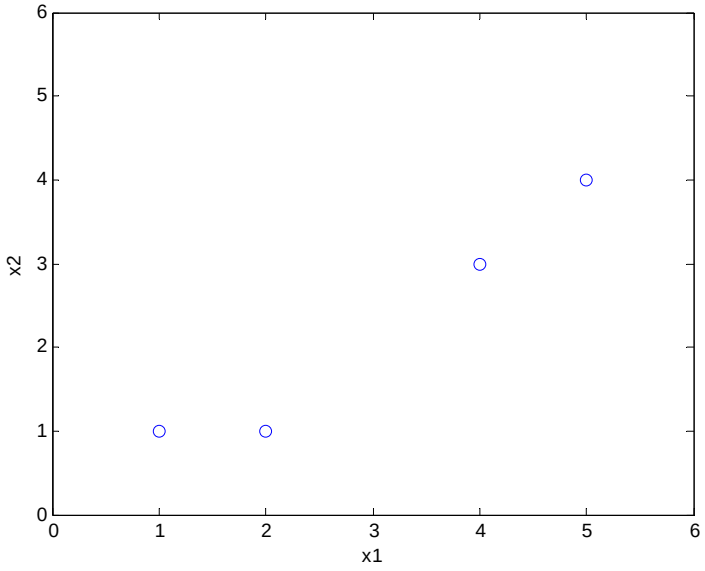
3a) [15 pts] Cluster the data shown below

$$X = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \end{bmatrix} \right\}$$

using single link agglomerative hierarchical clustering and city block distance (sum of the absolute value of differences per dimension) as the distance measure.

Show each step of the clustering clearly.

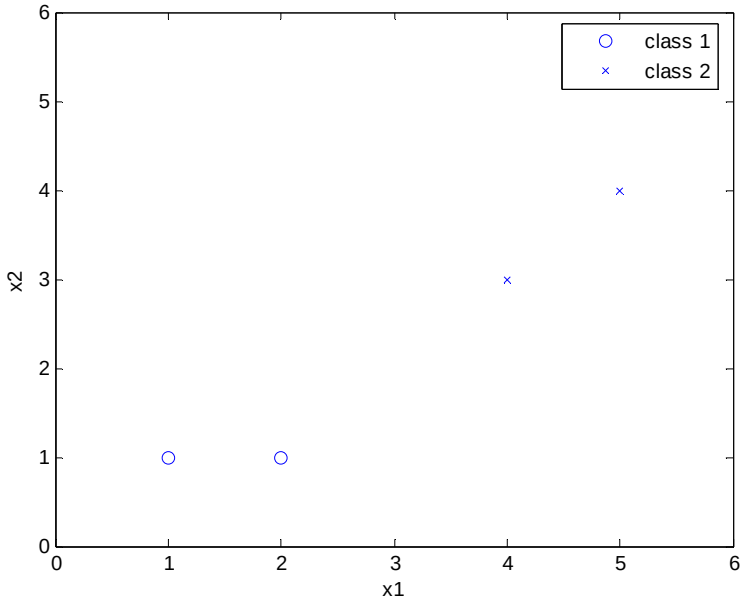
What are the coordinates of the cluster centers (means) if you want to obtain two clusters?



3b) [10 pts] Compare k-means and hierarchical clustering in terms of running time and performance? Suggest ways to improve performance of k-means clustering.

4) Consider the labeled data points given as follows:

$$X = \left\{ \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, 1 \right), \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, 1 \right), \left(\begin{bmatrix} 4 \\ 3 \end{bmatrix}, 2 \right), \left(\begin{bmatrix} 5 \\ 4 \end{bmatrix}, 2 \right) \right\}$$



Assuming that inputs are normally distributed with class covariance matrices are as follows:

$$S_1 = S_2 = s^2 I = s^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

4a) [20 pts] Compute the discriminant functions for both classes, $g_1(\underline{x})$ and $g_2(\underline{x})$.

4b) [5 pts] Compute and draw the discriminant function that separates the two classes.

Hint1: If $\underline{\mathbf{x}} \sim \mathrm{N}_d(\underline{\boldsymbol{\mu}}, \Sigma)$, then the pdf for x is given by:

$$p(\underline{\mathbf{x}}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})^T \Sigma^{-1} (\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}) \right]$$

Hint2: Use the log likelihood for the discriminant function.

Hint3: $(1.5)^2 = 2.25$, $(2.5)^2 = 6.25$, $(3.5)^2 = 12.25$, $(4.5)^2 = 20.25$