Name and Student ID:

Machine Learning BLG527E, Nov 20, 2012, 110mins, 9:30am-11:20am, Midterm Exam.

1	2	3	4	5	6	Total

Name: **Number:**
Signature:
Duration: 100 minutes.
Write your answers neatly in the space provided for them. Write your name on each sheet.
Books and notes are closed. Good Luck!

QUESTIONS

QUESTION1) [20 points, 5 points each] Explain (use at most six sentences per question, you can use drawings, formulas, etc. also):

1a) bias and variance of an estimator

1b) comparison of feature selection and clustering

1c) comparison of PCA and LDA

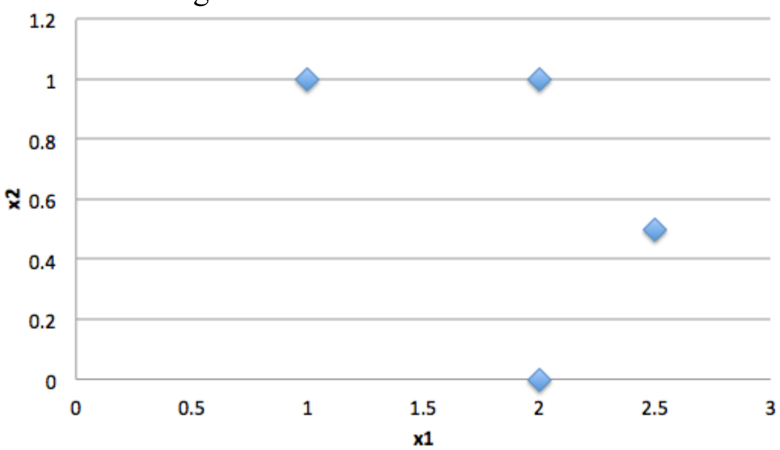
1d) comparison of Bayesian Decision Theory and parametric classification

Name and Student ID:

QUESTION2) [10 points] Two coins (Coin1 and Coin2) are flipped together 10 times and the following outcomes are observed. Do you think that Coin1 and Coin2's outcomes are independent of each other? Why or why not?

trial	1	2	3	4	5	6	7	8	9	10
Coin1	H	T	T	H	T	T	T	H	H	T
Coin2	T	H	H	T	H	H	T	T	T	H

QUESTION 3. [20 points]
Consider the four unlabeled data points
$\{x^1=(1,1), x^2= (2,1), x^3= (2.5,0.5), x^4= (2,0)\}$
shown in the figure:



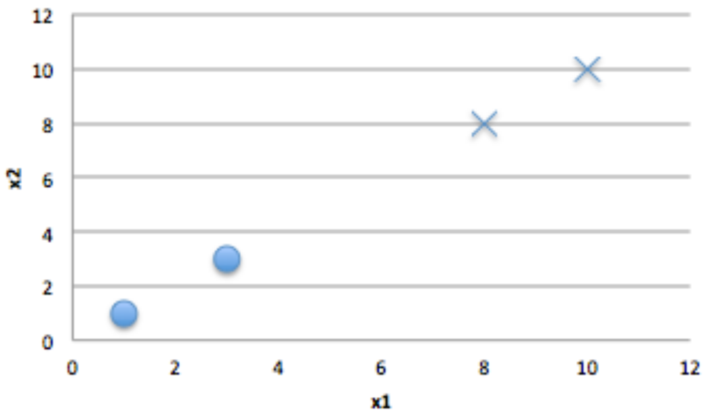
3a) [5 points]You need to divide them into 2 clusters. What would be the most reasonable clustering? Give the coordinates of the center of each cluster. (Use city block distance as the distance measure: $|a,b|=|a_1-b_1|+|a_2-b_2|$)

3b) [10 points]Considering all possible initial cluster centers (assume that 2 different cluster centers are chosen randomly among the four data points), what are the clusterings produced by the k-means clustering algorithm with k=2.

3c) [5 points]What are the clusters that would be produced by EM algorithm using mixtures of 2 Gaussians? (Give a sketch of how EM would work.)

Name and Student ID:

QUESTION 4. [25 points]
Consider the four labeled data points belonging to class O (class1) and class X (class 2) $\{x^1=(1,1), x^2=(3,3), x^3=(8,8), x^4=(10,10)\}$ as shown in the figure:



Assume that inputs in each class are normally distributed with the same class covariance matrices.

4a) [15 pts] Compute the discriminant function g(x) that separates these two classes. (For ease of computation, compute the actual covariance matrices and then take the diagonal elements to be a and b and the offdiagonals to be zero).

4b) [5 pts] Assume that you observed 8 more instances of class X, 4 of them at location $x^3=(8,1)$, and 4 of them at location $x^4=(10,1)$. Where would the discriminant be now?

4b) [5 pts] Assume that the data points were given as $\{x^1=(1,1), x^2=(3,1), x^3=(8,1), x^4=(10,1)\}$. Could you still compute the discriminant in 4a)? What method could you use to compute the discriminant?

Name and Student ID:

QUESTION4 Continued) Hints:

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

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

$$\det\left(\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}\right) = ad$$

Common Covariance Matrix S

- Shared common sample covariance **S**

$$\mathbf{S} = \sum_i \hat{P}(C_i)\mathbf{S}_i$$

- Discriminant reduces to

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}_i) + \log \hat{P}(C_i)$$

which is a linear discriminant

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

where

$$\mathbf{w}_i = \mathbf{S}^{-1}\mathbf{m}_i \quad w_{i0} = -\frac{1}{2}\mathbf{m}_i^T \mathbf{S}^{-1}\mathbf{m}_i + \log \hat{P}(C_i)$$

Name and Student ID:

QUESTION 5) [10pts] If you used PCA to transform all the 4 instances in Q4a) into one dimension, where would (1,1) be projected?

QUESTION 6. [15pts]

Compute P[C|W] for the Bayes Network shown below.

(C: Cloudy, S: Sprinkler was on, R: It rained, W: Grass is wet, ~ opposite of an event, for example, ~W: Grass is not wet.)