Roll No.:_____

CS 557: STATISTICAL PATTERN RECOG & LEARNING
Sessional 1, Fall 2016

Total Marks:30   **Total Time: 60 Minutes**

NOTE:
    1. You are allowed to bring one A4 sheet of paper with you (printed or handwritten).
    2. Sharing of calculators and A4 sheet is STRICTLY NOT ALLOWED
    3. None of the questions below require lengthy or complex calculations. Use simple methods for solving.
    4. In case of any ambiguity make a reasonable assumption

<div align="center">Good luck!</div>

QUESTION 1   (Marks: 5+5+2+3)

We have two classes, where the likelihood of data is modelled by Gaussian distribution. Suppose you are given the following statistics with two attributes x_1 and x_2:

For class 1: $\mu_1 = (0,0)$, $\Sigma_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$

For class 2: $\mu_2 = (2,1)$, $\Sigma_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$

Part a
What do the contours of the two distributions look like. Draw them separately on two graphs.

Part b
Find the decision boundary between the two classes assuming that we use Maximum aposteriori classifier and both classes are equally likely. Show all working.

Part c
On a separate graph plot both means and the decision boundary. Clearly write the coordinates of the points that meet the graph at any of the coordinate axis.

Part d
What is the classification of the points (1,1), (0.5,0.5) and (-1,-1). (No marks without proper working.)

QUESTION 2   (Marks: 5+5+5)

We have three types of documents fiction, politics and sports. We check whether the three words, i.e., *player*, *game*, *winner* occur in a document or not. Here are some statistics that we gather:

- Word *player* occurs in 40% of fiction, 60% of politics and 80% of sports documents.
- Word *game* occurs in 20% of fiction, 70% of politics and 90% of sports documents
- Word *winner* occurs in 50% of fiction, 30% of politics and 70% of sports documents
- There are 50% fiction documents, 20% politics documents and 30% sports documents

PART a
What is the probability that we observe no occurrence of *player* but occurrences of *game* and *winner* in any document if we use **naive Bayes' assumption**. Write the formula you are using as well.

PART b
If we observe no occurrences of *player* and *game* but an occurrence of *winner* in a document then which category of documents does it belong to**? You have to use Naive Bayes' assumption and MAP classification**. Show all working.

PART c
Suppose we are allowed to modify the statistics related to the word winner found in documents related to politics. What should be the minimum probability P(word=*winner* | document=politics) for a document to be classified as politics when we observe all three words *winner*,*player* and *game* in it and the classification is done via **naive Bayes' and maximum likelihood classification**. Show all working.