Machine Learning

Final Examination

Maintain Honor Code

There are 5 problems. Total points 116.

Problem 1	:	**30 Points**
Problem 2	:	**30 Points**
Problem 3	:	**25 Points**
Problem 4	:	**25 Points**
Problem 5	:	**06 Points**

1. You can use the text-books by K Murphy and S. Theodoridis. If you use any other book mention it. You can use class hand-outs. However you should derive all your results.
2. No consultation allowed with anyone.
3. You can consult me for any needed clarification. You can use any of the recommended text-books. You can also use class hand-outs, assignments, and a calculator.
4. Answer all five problems. *Please insert a box around your answers for readability.*

Name *(Capital Letters):*

Sign:

Questions

1. (30 points) We are given data points which belong to two different classes: C_1 and C_2. Let the set of two-dimensional data points be

$$X = \{x_1, x_2, x_3, x_4\}$$

The data points x_1 and x_2, belong to the class C_1, and data points x_3 and x_4 belong to the class C_2 respectively, where

$$\begin{aligned} x_1 &= \begin{bmatrix} 1 & 1 \end{bmatrix}^T, \quad x_2 = \begin{bmatrix} 3 & 5 \end{bmatrix}^T \\ x_3 &= \begin{bmatrix} -1 & 1 \end{bmatrix}^T, \quad x_4 = \begin{bmatrix} 1 & -3 \end{bmatrix}^T \end{aligned}$$

Determine the direction of the best linear discriminating vector.

Hint: See one of your assignments. Write explicitly the *steps*, as shown in the assignment problem.

Working Space

Working Space

2. (30 points) **Question** This problem is on principal component analysis (PCA). The set of data points is $\{x_1, x_2, x_3\}$. These are

$$x_1 = \begin{bmatrix} -1 & 0 \end{bmatrix}^T, \quad x_2 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T, \quad x_3 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$$

Determine the principal component of these data points. Use only the dominant eigenvalue. Also determine the corresponding error in the approximation two different ways.

Hint: See one of your assignments. Write explicitly the *steps*, as shown in the assignment problem.

Working Space

Working Space

3. (25 points) There are two parts in the problem

   (a) (18 points) Let $y = (y_1, y_2)$, and $w_1, w_2 \geq 0$ be constants.

   $$\max f(y) = \sum_{j=1}^{2} w_j \ln y_j$$

   $$\text{Subject to:} \quad (y_1 + y_2) = 1$$

   where $y_1, y_2 > 0$.

   i. (12 points) What are the optimal values of the y_j's? Justify your answer.

   ii. (6 points) The global maximum of $f(y)$ occurs at what values of y_1, and y_2. What is it?

   (b) (7 points) Generalize the result in part (a). Let $y = (y_1, y_2, \dots, y_K)$, and $w_j \geq 0, 1 \leq j \leq K$ be constants.

   $$\max f(y) = \sum_{j=1}^{K} w_j \ln y_j$$

   $$\text{Subject to:} \quad \sum_{j=1}^{K} y_j = 1$$

   where $y_j > 0, 1 \leq j \leq K$. The global maximum of $f(y)$ occurs at what values of $y_j, 1 \leq j \leq K$. What is it?

   It is not necessary to show all the steps in this part of the problem. Simply write the expressions.

Working Space

Working Space

4. (25 points) This problem is on directed graphical model. Each random variable in the set $\{X_1, X_2, X_3, X_4, X_5, X_6\}$ is specified by a node in the directed graph. The links in the graph specify dependencies of the random variables.

   (a) (2 points) What kind of graph is it?

   (b) (10 points) Write explicitly, an expression for the joint distribution

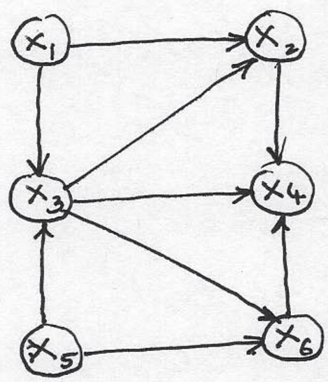
   $$P(X_1, X_2, X_3, X_4, X_5, X_6)$$

   in terms of conditional distributions $P(X_i \mid \Theta_i)$, where Θ_i is the set of dependent random variables of the random variable X_i, where $1 \le i \le 6$.

   (c) (13 points) Assume that each random variable X_i takes two values.

      i. (5 points) What is the size of the table to specify the joint distribution $P(X_1, X_2, X_3, X_4, X_5, X_6)$?

      ii. (8 points) What is the total size of the tables when tables of conditional distributions are used?



Working Space

5. (6 points) Name three well-known decision tree algorithms.