National University of Computer and Emerging Sciences, Lahore Campus

Course:	**Statistical Pattern Recog & Learning**	**Course Code:**	**CS 557**
Program:	**MS(Computer Science)**	**Semester:**	**Fall 2016**
Duration:	**60 Minutes**	**Total Marks:**	**20**
Paper Date:	**10-Nov-16**	**Weight**	**15%**
Section:	**ALL**	**Page(s):**	**4**
Exam:	**Midterm-2**	**Roll No:**	

Instruction/Notes:

1. This is an open book, open notes exam.
2. Sharing of calculators, books and notes is strictly not allowed.
3. In case of any ambiguity make a reasonable assumption.
4. Solve in the space provided. You can ask for extra sheets if needed.

Good Luck!

QUESTION 1 (Marks 5)

Suppose we are designing a system for regression on two input variables, x_1, x_2. Derive the normal equation for the weight w_0 if we use mean square error as the optimization criterion and the following expression for making predictions o for the output variable:

$$o = x_2 x_1 w_0 + x_1 w_1 + x_2^2 w_2$$

Roll Number: _____

QUESTION 2 **(Marks: 5)**

Suppose we have four points **w**,**x**,**y**,**z**. The following table shows the pairwise similarity of the 4 points. You have to apply agglomerative clustering using complete link clustering to this data and draw the final dendogram. No marks will be given if no working or iterations are shown.

	w	**x**	**y**	**z**
w	10	2	1	3
x	2	10	5	0
y	1	5	10	4
z	3	0	4	10

Roll Number: _____

QUESTION 3 **(Marks: 5)**

The following table shows the pairwise Euclidean distance of 5 points. All the 5 points (named **a**,**b**,**c**,**d**,**e**) lie in a 50 dimensional space. You have to find a mapping of all these points in 1 dimensional space that preserves the distances perfectly. Use any method you like. Show all working/reasoning.

	a	**b**	**c**	**d**	**e**
a	0	1	2	0	2
b	1	0	3	1	3
c	2	3	0	2	0
d	0	1	2	0	2
e	2	3	0	2	0

Roll Number: _____

QUESTION 4 (Marks 2+2+1)

Suppose we have the following data (one dimensional): 10,12,12,13,13,13,13,14,14,14

a. The Naive density estimate at x=13.2 for h=2 is _____

b. Suppose the density is given by (for some constant c) :

$$p(x) = \frac{1}{c}\sum_{t=1}^{N} k\left(\frac{x - x^{t}}{h}\right) \quad \text{and } k(x) = \begin{cases} 2 \; if \; |x| = 0 \\ 1 \; if \; |x| < 3 \\ 0 \; otherwise \end{cases}$$

The density at x = 12.2 for h=2 is _____

The density at x = 13.2 for h=1 is _____

c. If the label of the first 5 points is +1 and last five points is -1, then what is the classification of the point 12.6 when using 5 nearest neighbor algorithm.