


**National University of Computer and Emerging Sciences, Lahore Campus**

	<b>Course Name:</b>	Statistical and Mathematical Methods for Data Science	<b>Course Code:</b>	DS 501
	<b>Program:</b>	MS Data Science	<b>Semester:</b>	Fall 2019
	<b>Duration:</b>	60 Minutes	<b>Total Marks:</b>	30
	<b>Paper Date:</b>	November 07, 2019	<b>Weight</b>	
	<b>Section:</b>	N/A	<b>Page(s):</b>	2
	<b>Exam Type:</b>	Midterm Exam 2		

**Student : Name:** \_\_\_\_\_ **Roll No.** \_\_\_\_\_

- Instruction/Notes:**
1. Solve in the space provided. Extra sheets are NOT allowed
  2. One A4 handwritten help sheet is allowed.
  5. Sharing calculators is NOT allowed
  6. In case of any ambiguity make a reasonable assumption.

Good luck!

**PROBLEM 1 (Marks: 5)**

1. Find the eigen vectors and eigen values of the given matrix. Show all working:

$$\begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}$$

**PROBLEM 2 (Marks: 2+2+2+2+2)**

- 2a. Write (4,5) as a linear combination of vectors in the set: {(2,0),(1,1)}. How many combinations are possible?

- 2b. Add vectors to the set {(1,1,1),(1,2,1)}, so that it forms a basis for
- $\mathbb{R}^3$
- .

- 2c. What is the span{(1,1,0,0), (1,2,0,0)}. Is the span a vector space? Give reason.

- 2d. The regression co-efficients for a 2 variable problem are given by (10,5,-1). Here 10 is the intercept. What is prediction for the point (2,5). (show working)

- 2e. What would the regression co-efficients be close to if the shrinkage/regularization constant is set at infinity.

**PROBLEM 3 (Marks: 2.5+2.5)**

We have a Gaussian distribution for positive class given by  $\mu_1 = (3, 2)$  and  $\Sigma_1 = I$ . For the negative class, the Gaussian has  $\mu_2 = (-3, -2)$  and  $\Sigma_2 = I$ . Draw the contours of the two Gaussian distributions. Suppose we build a classifier using equal priors and using the two Gaussians for computing the likelihood. Plot the decision boundary and show the predicted classes for each region.

**PROBLEM 4 (Marks: 5)**

How many minimum test samples do you require to guarantee that the accuracy (ratio of total correct to total samples) is estimated with an error of at the most 0.05 with a 90% confidence. Show formula and working.

$$z_{0.10} = 1.282, z_{0.05} = 1.645, z_{0.025} = 1.960, z_{0.01} = 2.326, z_{0.005} = 2.576$$

**PROBLEM 5 (Marks: 5)**

The average time for running an SVM on 50 test samples was 20ms with a variance 2.5ms and the average time taken by naive Bayes' was 10ms with a variance of 4ms on 100 test samples. Can we claim that the running time for both algorithms is the same? Show formula and working. Justify your reasoning for proving or disproving this claim. (See problem 4 for z-values)