**MSc Data Science V Semester**

**Machine Learning**

**Problem Sheet – 1**

1. Write a program to solve the simultaneous equations:

X + Y =6

3X + 2y = 16

Plot the above equations and analyze the feasible solution.

1. Randomly create 100, 500, 1000 univariate data which follows the following distributions and find the maximum likelihood estimator.

Binomial distribution, Normal distribution, standard normal distribution and exponential distribution

1. Plot the following functions and check whether the following functions are positive definite or positive semi definite. Also check convexity or concavity for the same and find the relationship between concavity / convexity and positive / negative semi-definteness.
   1. Y = X2
   2. Y = X3
   3. Y = X
   4. Y = X Log X
   5. Y = |X|
2. Find the following metrics:
   1. Determinant
   2. Inverse matrix
   3. Rank
   4. Orthogonality
   5. Covariance
   6. Correlation

For the below matrices and infer the results

1. Unit
2. Square but not symmetric
3. Symmetric
4. Rectangular
5. Identity
6. Download the Iris data set from UCI machine learning repository - <https://archive.ics.uci.edu/ml/datasets/Iris>. The data set contains 4 input features and a class variable. Calculate the following :
   1. Find the rank of the data set. Check whether all features are linearly independent. Also find Nullity of the dataset.
   2. For each feature Find all statistical measures mean, standard deviation, covariance, correlation, skewness, kurtosis and interpret your results.
   3. For each feature X, find mean and subtract each value of feature from its mean
   4. draw box plot for each feature. Find outliers if any.
   5. Plot the graph for the following functions:
      1. Linear (y=x)
      2. Step function
      3. Sigmoid function
      4. RELU
      5. Modified RELU
      6. GELU
      7. tanh

For each pair of features

* 1. Draw scatterplot, give different color for each class, check the correlation between these features and check the possibility of linear classifier model
  2. Calculate Covariance and correlation matrices and infer the relationship between the corresponding features.
  3. Analyse the symmetricity and diagonal values in both matrices.
  4. Are these matrices orthogonal
  5. Calculate rank, determinant of thest matrices and infer the dependancy relations
  6. Calculate the eigen values and eigen vectors
  7. Find the L2 Norm of the above Eigen vectors
  8. Check whether these eigen vectors are unit vectors.
  9. Check these eigen vectors are orthogonal
  10. Find the Sum of Eigen values of covariance and correlation matrices and compare them with variance and correlation values respectively.
  11. Split the data into train and test in the ratio of 2:1, apply all machine learning algorithms, display the models (graphical representations) , predict the test set and print performance measures (Accuracy, precision, recall, F-measure, ROC, AUC).