Machine Learning (LAB 02)
Feature Engineering/Preprocessing

Exercise Programs

1. Create a vector (array) of 1XN dimension representing N-dimensional feature vector of a sample. Write a program to compute the mean and variance of the elements present in the array.

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1} \quad \text{Sample Variance}$$

$$\sigma^2 = \frac{\Sigma (x - \mu)^2}{N} \quad \text{Population Variance}$$

σ^2 = variance
X_i = the value of the ith element
$\bar{X}$ = the mean of X
N = the number of elements

   Comment what the mean and variance of sample represents.

2. Create two vectors each of dimension 1XM each representing N-dimensional feature vector of a sample. Write a program to compute the Covariance between them.

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

   Comment what Covariance between two vectors represents.

3. Create two vectors each of dimension 1XN. Write a program to compute the Correlation between them.

$$\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

   Comment what the Correlation represents.

4. Create a Matrix of MXN dimension representing the M-dimensional feature vector for N number of samples i. e $(i,j)^{th}$ entry of the matrix represents the ith feature of jth sample. Write a program to compute the covariance matrix and correlation matrix. Comment on takeaways from these matrixes.