Question 4 (SVM) -

a) When SVM is trained on US postal service dataset and using **linear kernel**:

- The Linear kernel is the simplest kernel function. It is given by the inner product <x,y> plus an optional constant c

$$k(x, y) = x^T y + c$$

- **Test Accuracy - 97.8773**
- **Number of Support Vectors - [12,12]**

b) Using linear kernel if we train the model on 50,100,200,800 rows the test error will be :

- **Trained using first 50 points-**
  Test accuracy - 96.933 %
  Number of support vectors - [2,2]
- **Trained using first 100 points-**
  Test accuracy - 98.349 %
  Number of support vectors - [3,3]
- **Trained using first 200 points-**
  Test accuracy - 97.405 %
  Number of support vectors - [4,4]
- **Trained using first 800 points-**
  Test accuracy - 97.877 %
  Number of support vectors - [10,10]

c) Now we are using polynomial Kernel and comparing its result at degree 2 and 5

i) False

- Because at C = 0.0001 and degree=2, training error = **2.24215 %** and at degree=5, training error = **0.6406149 %**. So clearly at degree=5 training error is less.

ii) True

- Because at C = 0.001 and degree=2, number of support vectors=**[76,76]** and at degree =5, number of support vectors=**[14,14]**. So clearly at degree=5 support vectors are less.

iii) False

- Because at C = 0.01 and degree=2, training error =**0.4484304 %** and at degree=5, training error = **0.4484304 %**. So clearly at degree=5 and degree=2 training error is same.

iv) True

- Because at C = 1 and degree=2, test error =**1.886792 %** and at degree=5, test error =**1.6509433 %** . So clearly at degree=5 test error is lower.

d) Now we are using rbf kernel and comparing its result at different different C

- **Training Error Comparison**

C	Training error(%)
C=0.01	0.384368
C=1	0.448430
C=100	0.320307
C=10000	0.256245
C=1000000	**0.128122**

So, training error at C=1000000 is lower.
When C is high margin in svm is low i.e you are overfitting the data.

- **Test Error Comparison**

C	Test error(%)
C=0.01	2.12264150
C=1	2.12264150
C=100	**1.88679245**
C=10000	**1.88679245**
C=1000000	2.12264150

So, testing error at C=100 and C=10000 is equal and lower.
For the same data training error is low at highest C but test error is low at C=100,10000