Programming Assignment 3

CLASSIFICATION AND REGRESSION

Team Number 25: **Course Number: CSE 574**
SHIVA MARISWAMY SUBRAMANI - 50133425
ANANDA KANAKARAJ SANKAR - 50133315
KARTHIK JANARTHANAN - 50133543

Logistic Regression:

Logistic regression is a type of probabilistic statistical classification model. It is used in estimating empirical values of the parameters in a qualitative response model.
Logistic regression can be binomial or multinomial. Binomial or binary logistic regression deals with situations in which the observed outcome for a dependent variable can have only two possible types (for example, true /false). Multinomial logistic regression deals with situations where the outcome can have more than two possible types.

Results:

Testing Accuracy: 91.992

Validation Accuracy: 91.55

Training Accuracy: 91.76

SUPPORT VECTOR MACHINE

Support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, and other tasks. For purpose of calculating performance metrics, good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.
SVM model and compute accuracy of prediction with respect to training data, validation data and testing using the following parameters:

1. Using linear kernel (all other parameters are kept default).
2. Using radial basis function with value of gamma setting to 1 (all other parameters are kept default).
3. Using radial basis function with value of gamma setting to default (all other parameters are kept default).
4. Using radial basis function with value of gamma setting to default and varying value of C.

LINEAR KERNEL:

Linear SVM will be very useful if original data being tested on is highly dimensional and original data is highly informative. Handwriting Classification Data is suitable for linear kernel as it is highly dimentional.

Training Set Accuracy	97.286
Validation Set Accuracy	93.64
Testing Set Accuracy	93.78

RADIAL BASIS FUNCTION:

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. The gamma parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors

(GAMMA = 1.0):

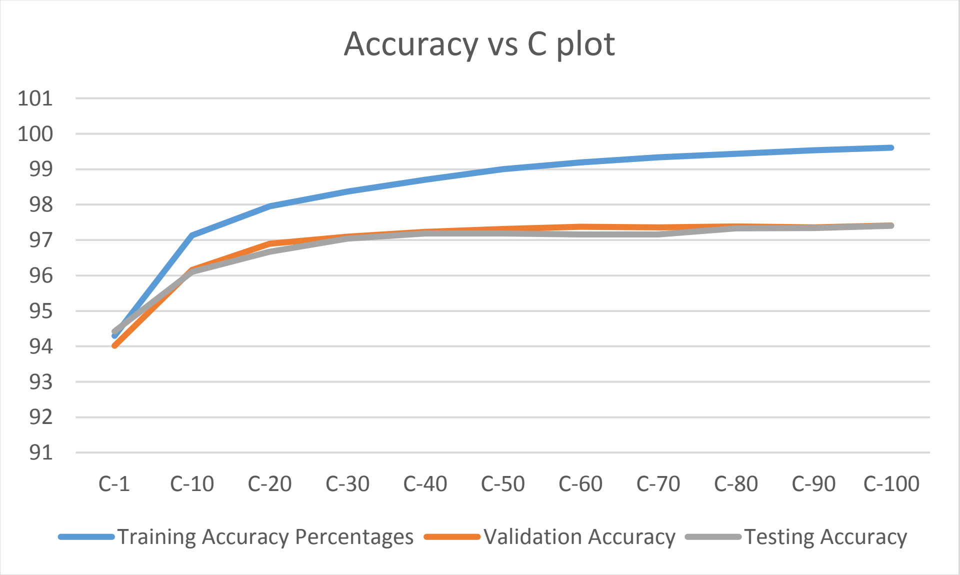
Training Set Accuracy	100
Validation Set Accuracy	15.48
Testing Set Accuracy	17.12

RADIAL BASIS FUNCTION (GAMMA = DEFAULT):

Training Set Accuracy	94.284
Validation Set Accuracy	94.0
Testing Set Accuracy	94.42

Varying C Values:

The C parameter tells the SVM optimization how much you want to avoid misclassifying each training example. For large values of C, the optimization will choose a smaller-margin hyper plane if that hyper plane does a better job of getting all the training points classified correctly. Conversely, a very small value of C will cause the optimizer to look for a larger-margin separating hyper plane.



C=1

Training Set Accuracy	94.294
Validation Set Accuracy	94.02

Testing Set Accuracy	94.42

C=10

Training Set Accuracy	97.132
Validation Set Accuracy	96.16
Testing Set Accuracy	96.1

C=20

Training Set Accuracy	97.952
Validation Set Accuracy	96.9
Testing Set Accuracy	96.67

C=30

Training Set Accuracy	98.372
Validation Set Accuracy	97.10
Testing Set Accuracy	97.04

C=40

Training Set Accuracy	98.706
Validation Set Accuracy	97.23
Testing Set Accuracy	97.19

C=50

Training Set Accuracy	99.002
Validation Set Accuracy	97.31
Testing Set Accuracy	97.19

C=60

Training Set Accuracy	99.196
Validation Set Accuracy	97.38
Testing Set Accuracy	97.16

C=70

Training Set Accuracy	99.34
Validation Set Accuracy	97.36
Testing Set Accuracy	97.16

C=80

Training Set Accuracy	99.438
Validation Set Accuracy	97.39
Testing Set Accuracy	97.33

C=90

Training Set Accuracy	99. 542
Validation Set Accuracy	97.36
Testing Set Accuracy	97.34

C=100

Training Set Accuracy	99. 616
Validation Set Accuracy	97.41
Testing Set Accuracy	97.40