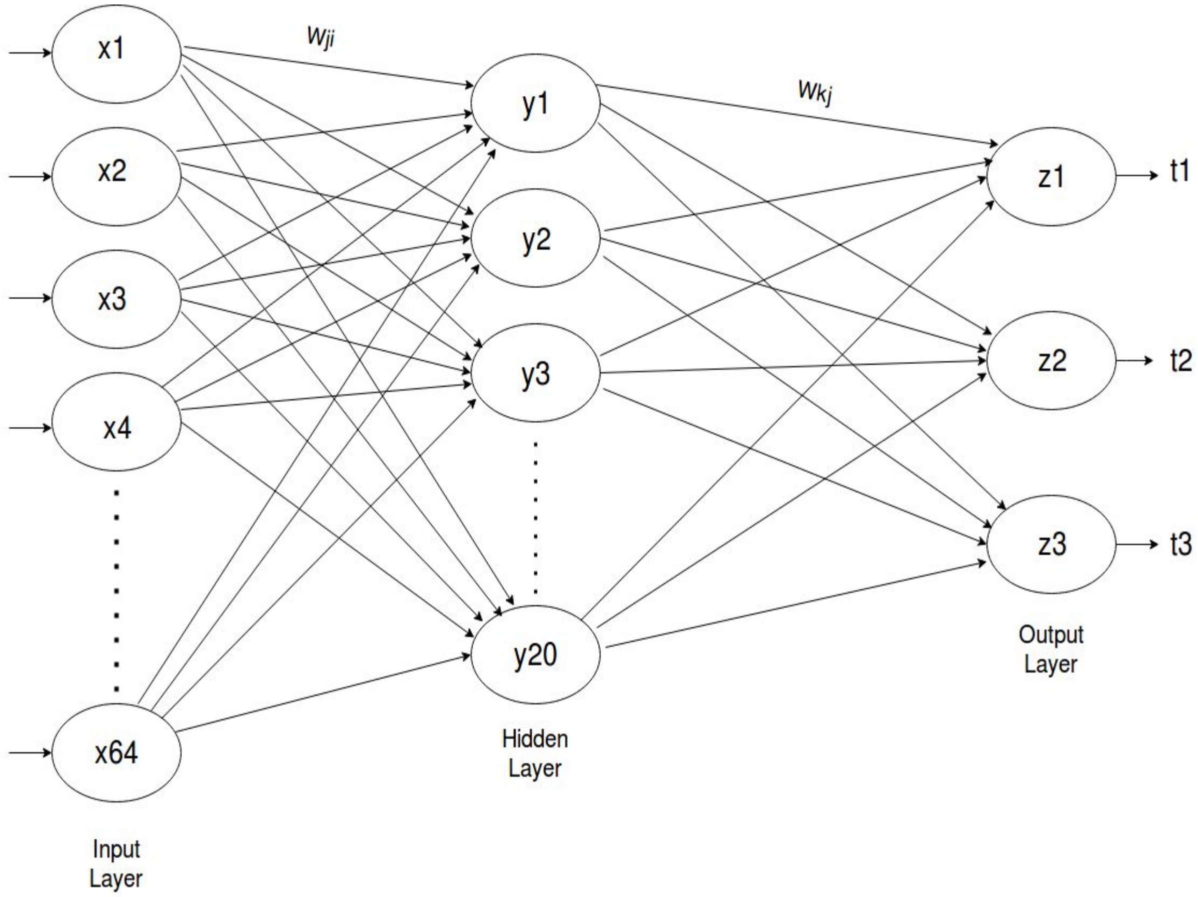
Assignment 4
CSE471 : SMAI

Name: Suryansh Agnihotri
Roll No. 20162037

1. Neural Network:

A. Architecture:



The program was tested for hidden layer(n_H) = 10,20 and 70. I have classified for all the 3 and 10 digits so the output layer has 10 nodes. The input layer is of 64 nodes in which the 32*32 matrix is downsampled to 8*8 matrix and sent as input.

B. Report:

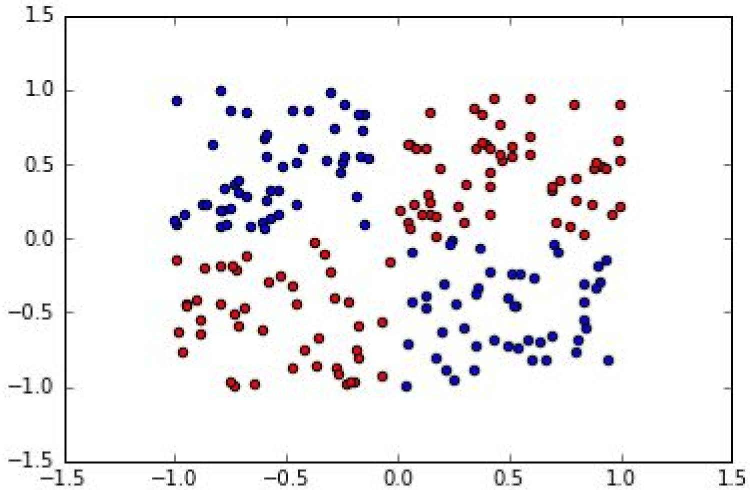
```
suryansh Desktop $ python neuralnet.py
0
0.351782955655
10
0.179524105673
20
0.179499385417
30
0.179557568252
40
0.179467106934
50
0.17937314761
60
0.17939047744
70
0.178837569321
80
0.176077230885
90
0.171250124486
100
0.166526112903
110
0.162599966332
120
0.159010601449
130
0.15532413472
140
0.151431891096
150
0.147839502176
160
0.145362674816
170
0.144215206985
180
0.143126267916
```

```
suryansh Desktop $ python neuralnet.py
0
0.206136240252
10
0.175355224455
20
0.168600525179
30
0.155477574722
40
0.135918386864
50
0.115268753692
60
0.0978164193101
70
0.0844154277186
80
0.0743269361976
90
0.0665932833974
100
0.0604799793528
110
0.0555049367563
120
0.051362388025
130
0.0478527049946
140
0.0448388220832
150
0.0422222584952
160
0.0399298287138
170
0.0379057671486
180
```

The program was tested for 200 epochs and after each 10 epochs the error was noted. In case of n_H=10(Fig 1) the error starts increasing after 100 epochs because of overfitting but in case of n_H=70(Fig 2) the value keeps on decreasing hence it is a better choice of n_H.

2. SVM:

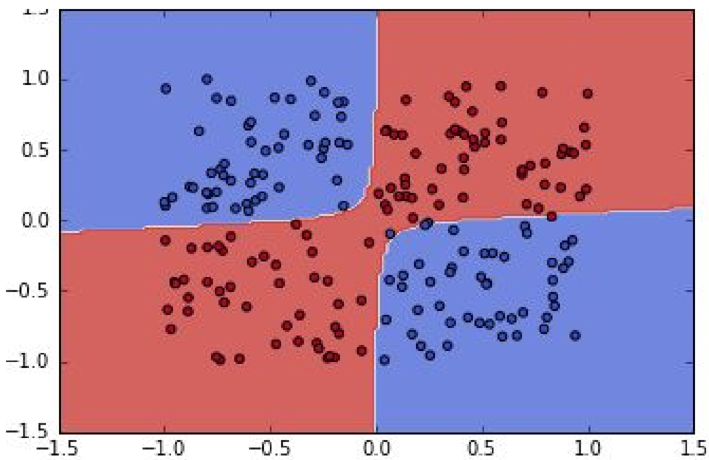
a. Polynomial Kernel:



Penalty parameter C of the error term.
kernel : string, optional (default='rbf') Specifies the kernel type to be used in the algorithm. It must be one of 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed' or a callable. If none is given, 'rbf' will be used.
degree : int, optional (default=3) Degree of the polynomial kernel function ('poly'). Ignored by all other kernels.

```
1 2
Accuracy: 0.97 (+/- 0.00)
1 3
Accuracy: 0.70 (+/- 0.55)
1 4
Accuracy: 0.74 (+/- 0.47)
10 2
Accuracy: 0.80 (+/- 0.45)
10 3
Accuracy: 0.72 (+/- 0.51)
10 4
Accuracy: 0.75 (+/- 0.48)
100 2
Accuracy: 0.78 (+/- 0.48)
100 3
Accuracy: 0.74 (+/- 0.50)
100 4
Accuracy: 0.76 (+/- 0.49)
1000 2
Accuracy: 0.79 (+/- 0.49)
1000 3
Accuracy: 0.76 (+/- 0.50)
1000 4
Accuracy: 0.77 (+/- 0.49)
```

Plotting boundary of polynomial kernel for c=1 and d=2



Gaussian Kernel

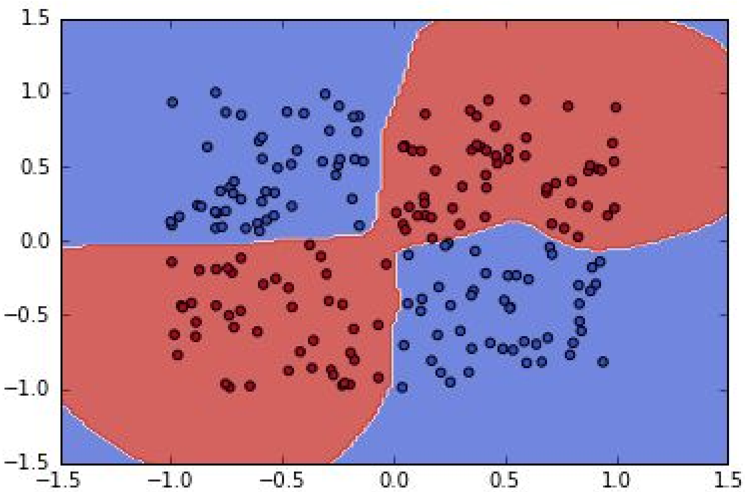
Intuitively, the gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. The gamma parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors.

The C parameter trades off misclassification of training examples against simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly by giving the model freedom to select more samples as support vectors.

```
0.01 0.1
Accuracy: 0.46 (+/- 0.00)
0.01 1
Accuracy: 0.46 (+/- 0.00)
0.01 10.0
Accuracy: 0.46 (+/- 0.00)
1 0.1
Accuracy: 0.59 (+/- 0.46)
1 1
Accuracy: 0.67 (+/- 0.52)
1 10.0
Accuracy: 0.73 (+/- 0.53)
100.0 0.1
Accuracy: 0.76 (+/- 0.53)
100.0 1
Accuracy: 0.79 (+/- 0.52)
100.0 10.0
Accuracy: 0.81 (+/- 0.50)
```

plotting gaussian kernel decision boundary for c=100 and gamma=10

```
Out[8]: (-1.5, 1.5)
```



c. Comparison:

Polynomial kernel: $K(X,Y)=(\gamma \cdot XTY+r)d,\gamma>0$
Radial basis function (RBF) Kernel: $K(X,Y)=\exp(\| X-Y \| \; 2 /2\sigma$
$2)$
which in simple form can be written as $\exp(-\gamma \cdot \| X-Y \| \; 2),\gamma>0$

From the above given images, it can be concluded that the mean accuracy of RBF is relatively better than polynomial kernel for different values of d.

3. Bayes Decision Theory:

a. Missing Value Handling:

Missing values are those for which values are not specified (in this dataset, these values are specified with ‘?’). Ignoring these values reduces the size of dataset and In practice, at prediction time, discarding instances with missing feature values may be inappropriate when the missing value has significant effect in deciding the classifier . So , here, a probabilistic approach is taken. I have assumed missing value has more chance to have that value for which the frequency is highest i.e. if value ‘v’ has occurred for ‘n’ times in the dataset of size ‘l’, then probability of any unknown value will be equal to ‘v’ is n/l, since n is the highest frequency, unknown value will have more probability to have the value ‘v’. So, in this case, Value is replaced with mode of the attribute.