Programming Assignment 1

Rachit Garg `CS14B050`

1 Feature Extraction

Different kernels were considered to fit the SVM on the DS2 dataset. Python libraries were used and we basically changed the hyper parameters to obtain good results after doing n-fold cross validation.
The strategy followed on a broad scale was to first have a big step function and find values from a larger range around which a smaller step function was chosen and a finer analysis was done.

1.1 Polynomial Kernel

We change the parameters such as the constant coefficient the degree and the slack variable, we run this through loops and note the outcome.

- Best degree, coefficient noted are $Degree : 6$, $coefficient : 1.25$.
- Accuracy, precision, recall, f-measure were noted and compared.
- Best performance results by Polynomial kernel are $Accuracy : 0.500$, $Precision : 0.5694$, $recall : 0.500$, $f-measure : 0.509513$.
- The relation for accuracy of training data vs degree is shown by the graph.

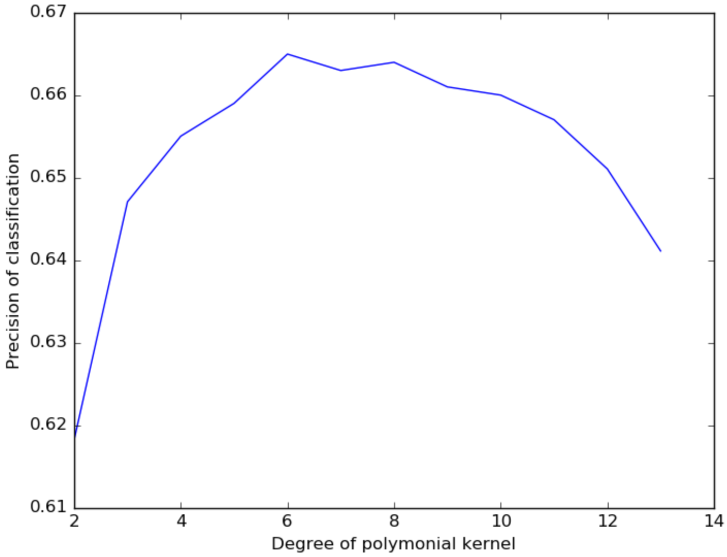

Figure 1: Precision(Cross validation on training set) vs degree

1.2 Sigmoid Kernel

We change the parameters such as the constant coefficient, gamma and note the accuracy .

- Best gamma value, coefficient noted are value: $Gamma : 0.08$, $coefficient : 0.2$.
- Accuracy, precision, recall, f-measure were noted and compared.
- Best performance results by Sigmoid kernel are $Accuracy : 0.6250$, $Precision : 0.65125$, $recall : 0.625$, $f - measure : 0.608025$.
- The relation for accuracy of training data vs degree is shown by the graph.

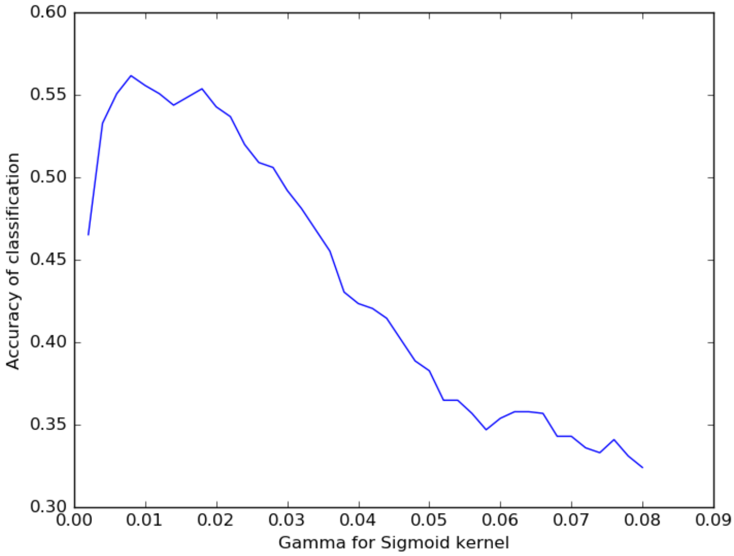

Figure 2: Accuracy(cross validation on training set) vs gamma

1.3 Linear Kernel

We change the parameters such as the slack in this kernel. .

- Value of C that gives best results are: $C : 1$.
- Accuracy, precision, recall, f-measure were noted and compared.
- Best performance results by Linear kernel are $Accuracy : 0.50$, $Precision : 0.526671$, $recall : 0.50$, $f - measure : 0.475493$.

1.4 Gaussian Kernel

Here the parameter we have to tune is gamma, a kernel parameter. Gamma comes in the negative exponent of kernel implying a higher gamma means points are dense in kernel space. The output we see is:

- Value of C ,gamma, that gives best results are: $C : 1.8$, $gamma : 0.0095$.
- Accuracy, precision, recall, f-measure were noted and compared.
- Best performance results by Gaussian kernel are $Accuracy : 0.25$, $Precision : 0.0625$, $recall : 0.25$, $f - measure : 0.10$.

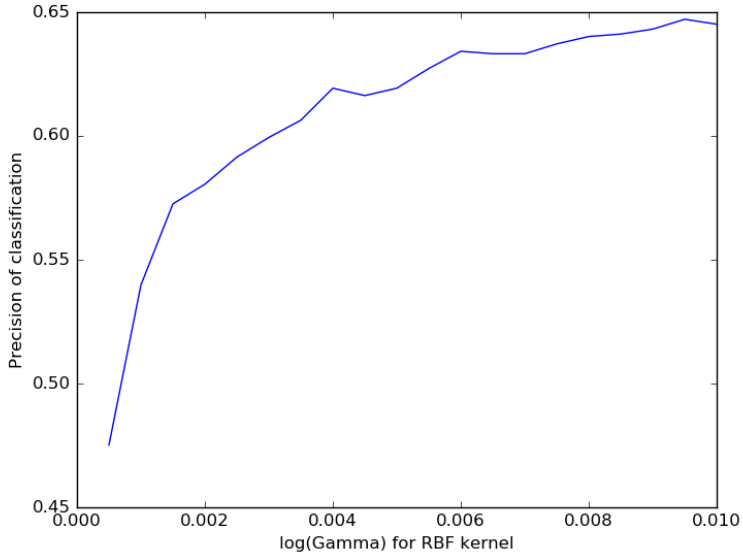

Figure 3: Accuracy(cross validation on training set) vs gamma

2 ANN - Backpropogation

2.1 Without regularization

ANN was chosen to classify the DS2 data set used in last assignment. The script that was attached with the last dataset itself was run.

- Least squares fit was implemented as an error function.
  $R(\theta) = \Sigma_k \Sigma_i (y_{ik} - f_k(x_i)^2$
- Centroids of the classes were close enough so that there is overlap in the classes.
- Softmax function was used as activation function in final layer, to prevent masking.
- Initial values of weights lie in [-0.7,0.7].
- Performance results by the ANN noted were,(epochs till 1000 were noted). Performance on test data: $Accuracy$: 0.5125, $Precision$: 0.580397, $recall$: 0.512500, $f-measure$: 0.492163.

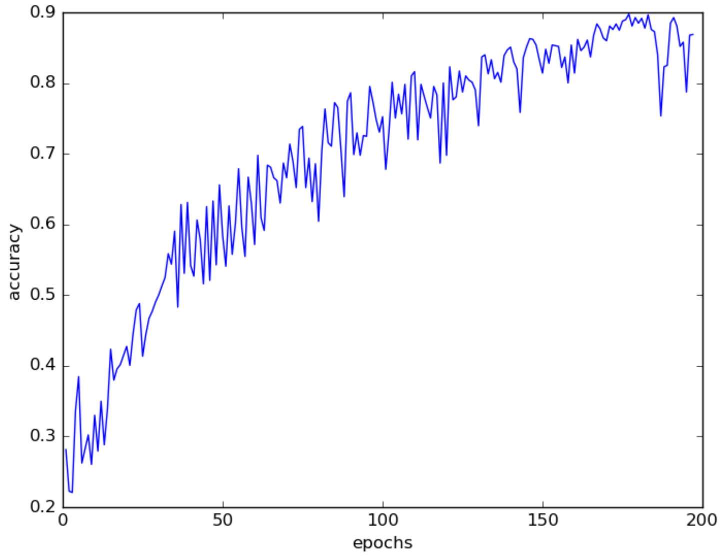

Figure 4: Accuracy(Of ANN without regularization on train set) vs epochs

2.2 With regularization

Regularized Least squares fit was implemented as an error function.
$R(\theta) = \Sigma_k \Sigma_i (y_{ik} - f_k(x_i))^2 + \lambda(\Sigma_k \Sigma_m \beta_{km}^2 + \Sigma_m \Sigma_l \alpha_{ml}^2)$
Gradient descent rule was updated with terms $2\lambda a$ and $2\lambda b$ added. Observations were as follows:
Performance results by the ANN noted were,(epochs till 1000 were noted or till accuracy improves above 0.9).

- Performance on test data with lambda - 0.01: $Accuracy: 0.5375$, $Precision: 0.606499$, $recall: 0.5375$, $f-measure: 0.531761$.
- Performance on test data with lambda - 0.1: $Accuracy: 0.6125$, $Precision: 0.658807$, $recall: 0.612500$, $f-measure: 0.611671$.
- Performance on test data with lambda - 1: $Accuracy: 0.4$, $Precision: 0.218154$, $recall: 0.4$, $f-measure: 0.269405$.
- Performance on test data with lambda - 10: $Accuracy: 0.25$, $Precision: 0.0625$, $recall: 0.25$, $f-measure: 0.100$.
- Performance on test data with lambda - 100: $Accuracy: NaN$, $Precision: Nan$, $recall: NaN$, $f-measure: Nan$.

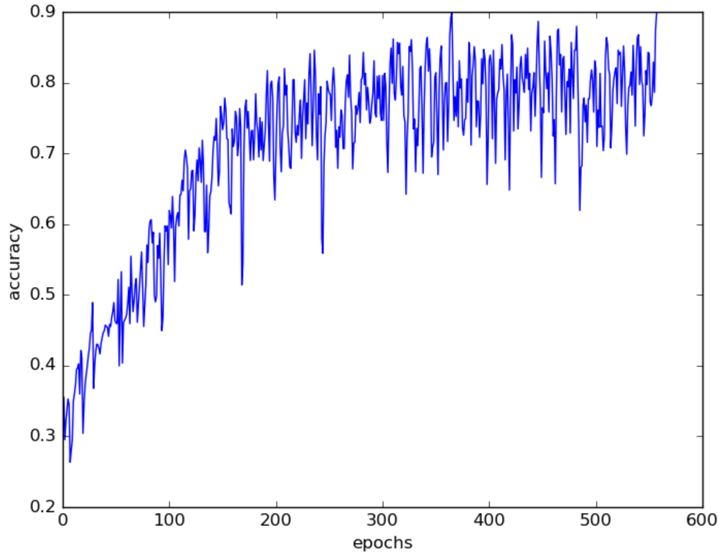

Figure 5: Accuracy(Of ANN with $\lambda = 0.01$ on train set) vs epochs

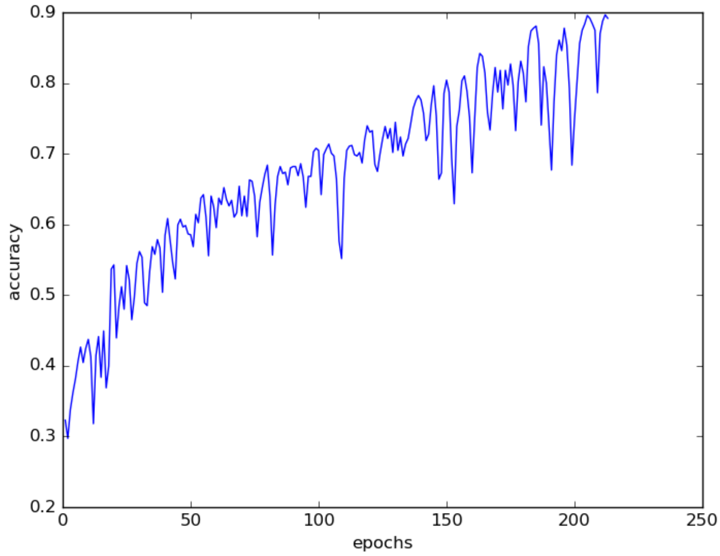

Figure 6: Accuracy(Of ANN with $\lambda = 0.1$ on train set) vs epochs

3 Inferences

- The best results of training and test error were observed when λ was 0.1. It was observed that for small values of λ slight over fitting takes place and for larger λ 's the weights are driven to zero resulting points to be classified to a class with probability 0. Due to which incorrect answers are produced.
- General stopping criteria was to stop when the training error goes below 10 percent.
- α has a higher growth rate than β, the parameter by which we gradient descent (Since it has a larger number of parameters).
- Learning rates do not lead to convergence in the last two cases, 'NaN's or random outputs are seen as weights are driven to zero and all is classified into one class.
- The classifier we learn is a strong classifier.

4 Weka Analysis on Mushroom data set

The following observations were made

- The value of MinNumObj is that the minimum instances per leaf guarantees that at each split, at least 2 of the branches (but not necessarily more than 2) will have the minimum number of instances.
  In our data set, changing the MinNumObj had no effect. Unless changed

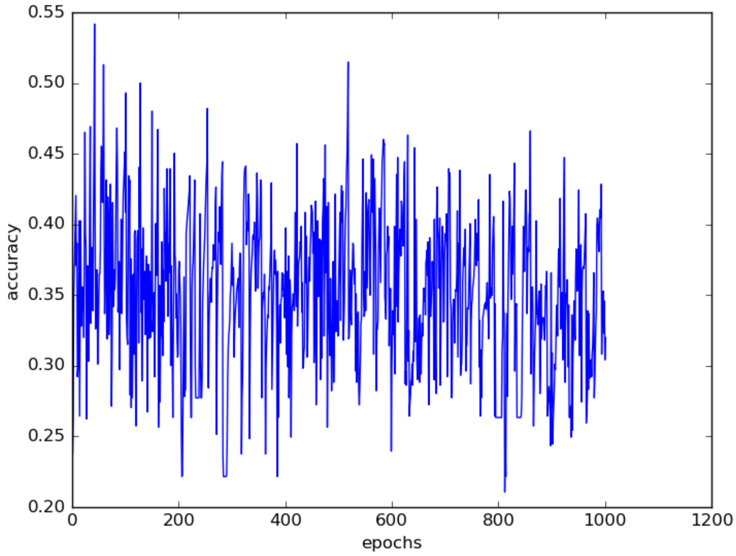

Figure 7: Accuracy(Of ANN with $\lambda = 1$ on train set) vs epochs

to very large values, which did not decrease node tree much and introduced an error of 1 percent.

- Pruning the data led to a much smaller and more interpretable tree.(Without any significant change in error)In some cases it helps prevent overfitting. In our data set, the size of tree on pruning and number of leaves remained same, however some split points around which split was made have changed.
- At a glance the important features seem to be - odor, stalk-shape, spore-print-color, gill-size, gill-spacing, population.