CS5011: Introduction to Machine Learning **Fall 2014**

Programming Assignment 2 Report

16/10/2014 *Ditty Mathew, CS13D018*

1 Question 1

1.1 Feature Extraction

The given dataset is of the form :
265 images as training set for category coast
239 images as training set for category forest
223 images as training set for category insidecity
279 images as training set for category mountain

And there are 20 images for each classes in the test set. We extracted the RGB features from these images using image histogram and re-bin each R,G,B of size 256 into 32 bins. So all together, each image is of size 1x96 after re-bining. Finally, normalized the data. The class labels for each class were set as 1, 2, 3, 4 respectively for classes coast, forest, insidecity, mountain.

1.2 SVM

We used 5-fold cross validation to find the best model parameters. Since each classes are having unequal number of instances for training, we splitted each class separately and finally concatenated all classes for each folds. Result obtained for each kernels are as follows,

1.2.1 Linear Kernel

C	Accuracy
1	59.38
100	58.97
1000	55.35
10000	55.14

Test Set Accuracy with C =1 = 56.25

1.2.2 Polynomial Kernel

C	Degree	Accuracy
1	1	30.05
1	2	31.96
1	3	29.73
100	1	59.28
100	2	48.39
100	3	29.73
1000	1	61.29
1000	2	62.00
1000	3	29.73
10000	1	58.77
10000	2	66.43
10000	3	56.96

Test set Accuracy (using C=10000 and degree=2) = 58.75

1.2.3 Gaussian Kernel

C	Gamma	Accuracy
1	0.1	59.57
1	0.01	45.97
1	0.001	30.15
1	1	65.22
1	10	49.39
1	100	28.02
100	0.1	63.81
100	0.01	62.40
100	0.001	56.45
100	1	63.11
100	10	51.11
100	100	28.52
1000	0.1	59.27
1000	0.001	60.50
1000	1	63.00
1000	10	51.11
1000	100	28.52
10000	0.1	58.87
10000	0.01	60.18
10000	0.001	61.30
10000	1	63.00
10000	10	51.11
10000	100	28.52

Test set, Accuracy (Using C= 1, Gamma=1) = 25

1.2.4 Sigmoid Kernel

C	Accuracy
1	26.71
C= 100	26.71
C= 1000	26.71
C= 10000	26.71

Test set, Accuracy (Using C= 1) = 25

Question 2

Implemented original back propagation algorithm. Parameters used are as follows
No of hidden layers = 1
No of Hidden Nodes = 40
Learning rate = 0.000001

Per-class precision, recall and f-measure obtained is as follows,

Class	Precision	Recall	F-measure
coast	0.41	0.8	0.54
forest	0.42	0.25	0.31
insidecity	0.36	0.40	0.38
mountain	0.43	0.15	0.22

Accuracy = 40%

1.3 Alternate Error Function

The derivation is as follows,

$$
\begin{aligned}
R(\theta) &= \frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{K}(y_{ik} - f_k(x_i))^2 + \gamma(\sum_{k}\sum_{m}\beta_{km}^2 + \sum_{m}\sum_{l}\alpha_{ml}^2) \\
Z_{mi} &= \sigma(\alpha_{0m} + \alpha_m^T x_i) \\
\\
Z_i &= (Z_{1i}, Z_{2i}, ..., Z_{mi}) \\
\frac{\partial R}{\partial \beta_{km}} &= -(y_{ik} - f_k(x_i))g_k^{'}(\beta_k^T Z_i)Z_{mi} + 2\gamma\beta_{km} \\
\frac{\partial R}{\partial \alpha_{ml}} &= -\sum(y_{ik} - f_k(x_i))g_k^{'}(\beta_k^T Z_i)\beta_{km}\sigma^{'}(\alpha_m^T x_i)x_{il} + 2\gamma\alpha_{ml} \\
\beta_{km}^{(r+1)} &= \beta_{km}^{(r)} - \gamma^{(r)}\Sigma_{i=1}^{N}\frac{\partial R}{\partial \beta_{km}^{(r)}} \qquad (1) \\
\alpha_{ml}^{(r+1)} &= \alpha_{ml}^{(r)} - \gamma^{(r)}\Sigma_{i=1}^{N}\frac{\partial R}{\partial \alpha_{ml}^{(r)}} \\
\frac{\partial R}{\partial \beta_{km}} &= \delta_{ki}Z_{mi} + 2\gamma\beta_{km} \\
\frac{\partial R}{\partial \alpha_{ml}} &= s_{mi}x_{il} + 2\gamma\alpha_{ml} \\
\\
s_{mi} &= \sigma'(\alpha_m^T x_i)\Sigma_{k=1}^{K}\beta_{km}\delta_{ki} \\
\delta_{ki} &= -(y_{ik} - f_k(x_i))g_k'(\beta_k^T Zi)
\end{aligned}
$$

Results

Parameters used are as follows
No of hidden layers = 1
No of Hidden Nodes = 60
Learning rate = 0.000005

Per-class precision, recall and f-measure obtained for $\gamma = 0.01$ is as follows,

Class	Precision	Recall	F-measure
coast	0.14	0.10	0.12
forest	0.28	0.65	0.39
insidecity	0.23	0.15	0.18
mountain	0.33	0.10	0.15

Accuracy = 25%

Per-class precision, recall and f-measure obtained for $\gamma = 0.1$ is as follows,

Class	Precision	Recall	F-measure
coast	0.83	0.25	0.38
forest	0.00	0.00	0.00
insidecity	0.28	0.95	0.44
mountain	0.00	0.00	0.00

Accuracy = 30%

Per-class precision, recall and f-measure obtained for $\gamma = 1$ is as follows,

Class	Precision	Recall	F-measure
coast	0.50	0.05	0.09
forest	0.00	0.00	0.00
insidecity	0.11	0.20	0.14
mountain	0.18	0.35	0.24

Accuracy = 15%

Per-class precision, recall and f-measure obtained for $\gamma = 1$ is as follows,

Class	Precision	Recall	F-measure
coast	0.25	0.80	0.38
forest	0.33	0.10	0.15
insidecity	0.00	0.00	0.00
mountain	0.50	0.05	0.09

Accuracy = 24%

Per-class precision, recall and f-measure obtained for $\gamma = 1$ is as follows,

Class	Precision	Recall	F-measure
coast	0.00	0.00	0.00
forest	0.00	0.00	0.00
insidecity	0.21	0.45	0.29
mountain	0.30	0.55	0.39

Accuracy = 25%

As the value of γ increases the weights tend to shrink to zero. This helps to overcome the problem of overfitting training data.

Question 3

The extracted features for the categories Forest and Mountain alone have been used for this question. And converted the class labels of each class to -1 and +1 respectively.

Logistic Regression

Per-class precision, recall and f-measure are reported as follows

class	precision	recall	f-measure
Forest	1.00	0.90	0.95
Mountain	0.91	1.00	0.95

1.4 L1-regularized Logistic Regression

Per-class precision, recall and f-measure are reported as follows

.

With lamda=0.1

class	precision	recall	f-measure
Forest	0.9	0.90	0.90
Mountain	0.9	0.90	0.90

.

With lamda=0.01

class	precision	recall	f-measure
Forest	1	0.95	0.97
Mountain	0.95	1	0.98

.

With lamda=0.001

class	precision	recall	f-measure
Forest	0.9	0.9	0.9
Mountain	0.9	0.9	0.9

.

With lamda=0.0001

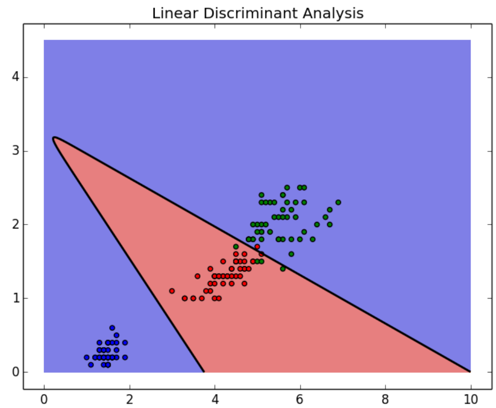
class	precision	recall	f-measure
Forest	0.86	0.90	0.88
Mountain	0.89	0.85	0.87

Question 4

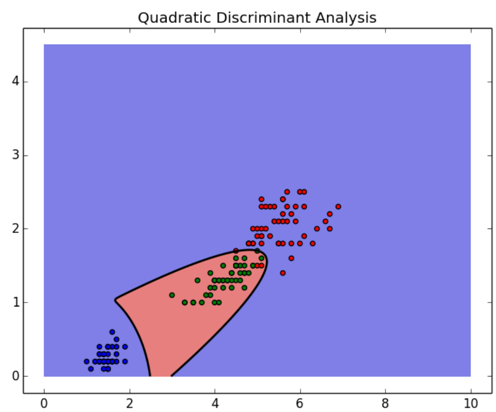
The Iris dataset consists of 3 classes of 50 instances each, where each class refers to a type of iris plant. The classes are Iris Setosa, Iris Versicolor, and Iris Virginica. The attributes are sepal length, sepal width, petal length and petal width, each in cm. Only the last two features are to be used for this experiment here.

We need to perform LDA, QDA and RDA(varying the lambda values) and visualize the boundaries learnt. The Plots obtained are shown below.

Visualization of classifier boundaries after Linear Discriminant Analysis has been performed:



Visualization of classifier boundaries after Quadratic Discriminant Analysis has been performed:

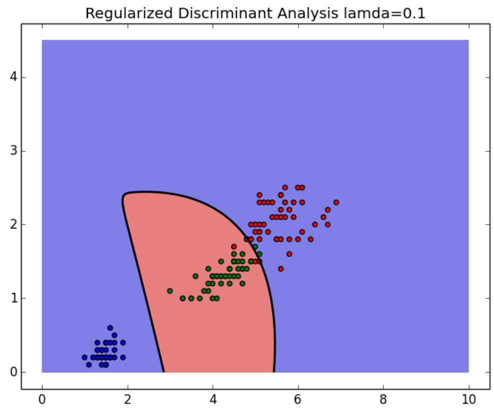


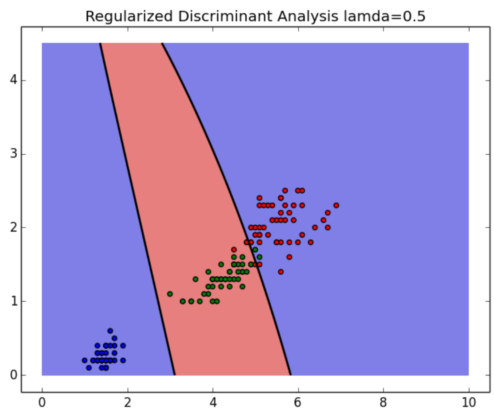
In RDA, the regularized covariance matrices have the form,

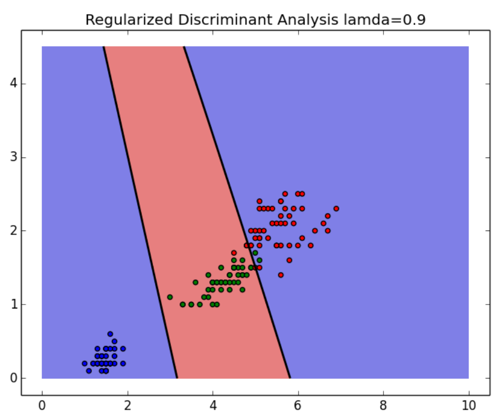
$$\hat{\Sigma_k}(\lambda) = \lambda\hat{\Sigma_k} + (1-\lambda)\hat{\Sigma_k} \tag{2}$$

Here $\lambda\epsilon[0,1]$.

Visualization of classifier boundaries after Regularized Discriminant Analysis with lamda =0.1,0.5,0.9 has been performed. And optimum got for lambda=0.9.


Regularized Discriminant Analysis lamda=0.1
0
1
2
3
4
0
2
4
6
8
10


Regularized Discriminant Analysis lamda=0.5
0
1
2
3
4
0
2
4
6
8
10


Regularized Discriminant Analysis lamda=0.9
0
1
2
3
4
0
2
4
6
8
10