Assignment 2

Shubham Singla, Harsh Kumar and Rupesh Kashyap

1 Introduction

FOLLOWING report contains observation for the 4 data sets provided in the assignment namely, FMNIST, Medical data, Railway data and River Data set. The observations contain the performance of different machine learning algorithms on the above mentioned data sets. The algorithms implemented are PCA, Linear regression with polynomial of different degrees phi functions, logitic regression using softmax, FLDA, PCA and SVM for multi class cases. Results have been summarized below in the form of accuracy, precision, recall, f-score, micro and macro average precision, recall and f-score.

2 Medical Data

Each of 'Healthy', 'Surgery' and 'Medication' are assigned a number 0, 1 and 2 respectively. Thus, it reduces to a problem of having labeled data with 3 classes and 3 features for each data point. The data was observed in 3D and 2D using PCA. Standardization of data followed by a 75/25 split into train and dev set was done. Perceptron, FLDA, SVM (with RBF, Linear, Polynomial and sigmoid kernel) and Logistic Regression (involving softmax activation) were applied. Perceptron and FLDA used a similar approach as of softmax for classification.

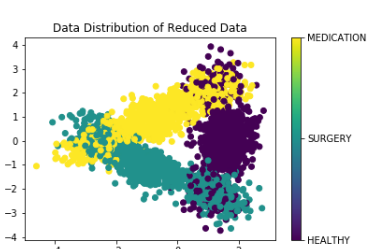


Fig. 1 Visualization of Data in 2D

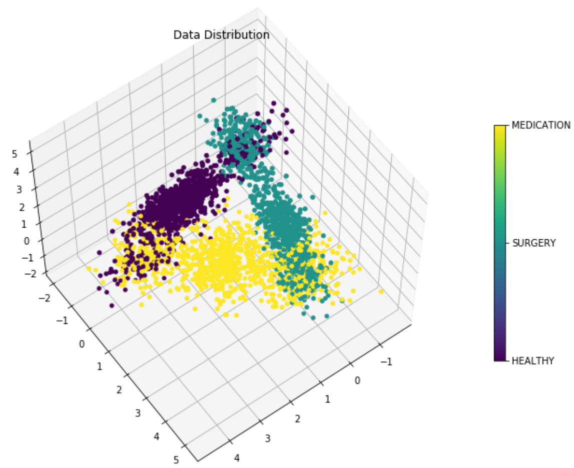


Fig. 2 Visualization of Data in 3D

2.1 Observations

2.1.1 Perceptron Model

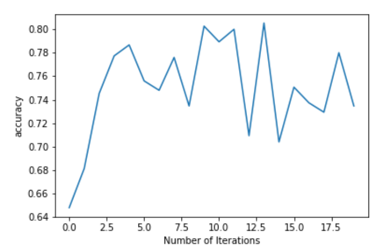


Fig. 3 Variation of Accuracy on Y-axis Vs Number of iterations*500 + 100 on X-axis

Perceptron model gave best accuracy after 6600 iterations. Performance on test data-
Accuracy - 0.8053
F-Score (Class 1) - 0.8523
F-Score (Class 2) - 0.8371
F-Score (Class 3) - 0.7
Macro Average F-Score - 0.8108
Micro Average F-Score - 0.8053

2.1.2 FLDA Model

Fisher Linear Discriminant Analysis gave the following results - Performance on test data-
Accuracy - 0.811
F-Score (Class 1) - 0.8738
F-Score (Class 2) - 0.7894
F-Score (Class 3) - 0.7728
Macro Average F-Score - 0.8139
Micro Average F-Score - 0.811

2.1.3 Logit Model (Softmax Function)

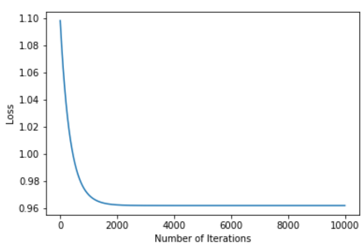


Fig. 3 Variation of Loss Vs Number of iterations

Loss in logit model stopped decreasing after 2000 iterations. Performance on test data-
Accuracy - 0.8053
F-Score (Class 1) - 0.8523
F-Score (Class 2) - 0.8371
F-Score (Class 3) - 0.7
Macro Average F-Score - 0.8108
Micro Average F-Score - 0.8053

2.1.4 SVM Model

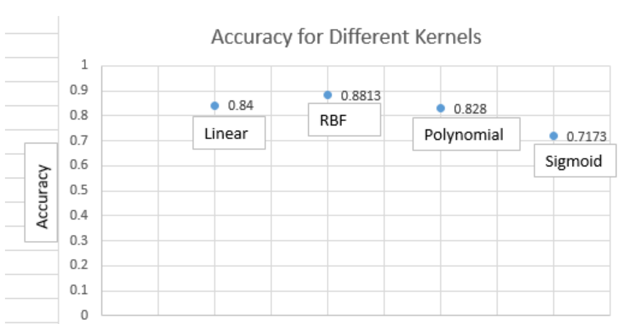


Fig. 3 Comparing Accuracy on Y-axis Vs Different kernels used on X-axis

Support Vector Machines gave best accuracy with rbf kernel. Performance on test data-

SVM Classifier

Kernel	Accuracy	F(C1)	F(C2)	F(C3)
Linear	0.84	0.9018	0.8366	0.7848
RBF	0.8813	0.9127	0.8893	0.8405
Poly	0.828	0.8447	.8288	0.8139
Sigmoid	0.7173	0.7504	0.7284	0.6667

2.2 Conclusion

Maximum accuracy of 0.88 for SVM classifier with rbf kernel is obtained. For logit model, loss decreased with the number of iterations and it stopped decreasing after a certain number of iterations. Perceptron model didn't converge meaning that the data is not linearly separable. Accuracy for perceptron model increased to a certain number of iterations after which it starts fluctuating. Maximum accuracy was obtained at around 6600 iterations. It was also observed that on applying PCA and reducing the data to 2 dimensions resulted in decrease in the accuracy.

3 River Data

River data contains a lot of noise as is clear from the image. It seems quite plausible that it won't be able to fit using linear model. Polynomial features were also tried and the results are tabulated below in the form of r-squared values.

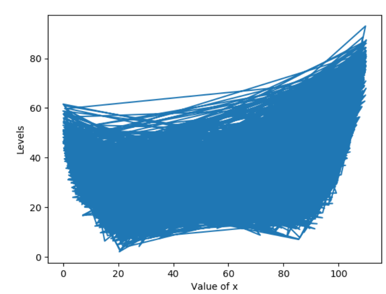


Fig. 6 Visualizing River levels with x

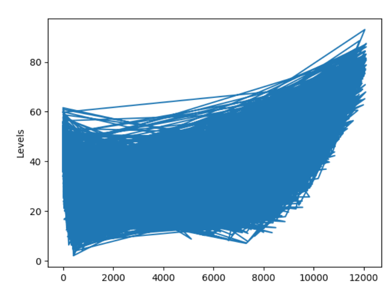


Fig. 7 Visualizing River levels with squared of x

3.1 Observations

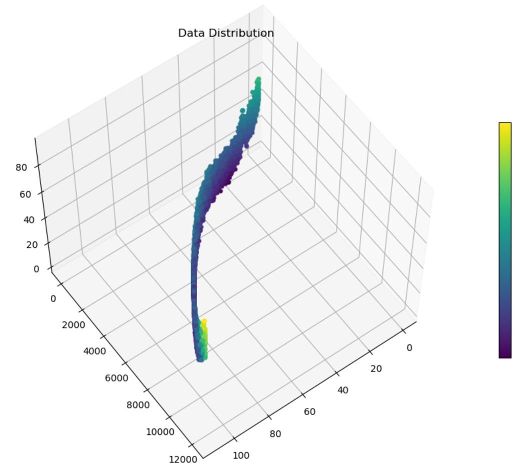


Fig. 7 Visualizing River levels with x and x-squared

Linear Regression

Degree of Polynomial	R-squared
One	0.126
Two	0.456
Three	0.498
Four	0.823
Five	0.823

Maximum R-squared is obtained for polynomial of **degree 4 - 0.823**.

3.2 Conclusion

We observed that value of R-squared became constant with increasing degree of polynomial. Also, we tried fitting exponential, square root and log after clipping minimum value to zero, results were more or less same.

4 Railway Data

The labels corresponding to each data point were in binary format denoting BOARDED AND NON-BOARDED CASES. This dataset is example of categorical data and discrete. The columns of SEX of passenger and TYPE OF COACH booked is converted into binaries using ONE-HOT encoding. Later models FLDA, perceptron, Logistic regression and SVM models are tested on standardised and non-standardised data.

1) Using PCA to visualize the data in 2 dimensions.

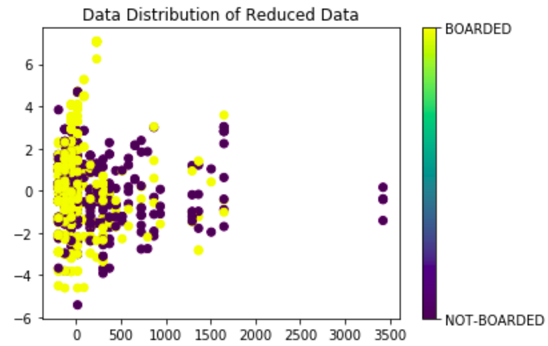


Fig. 1 Visualizing railway data in 2D using PCA

2) Visualizing the class counts as this kind of data is usually skewed and this may affect the performance of classifiers.

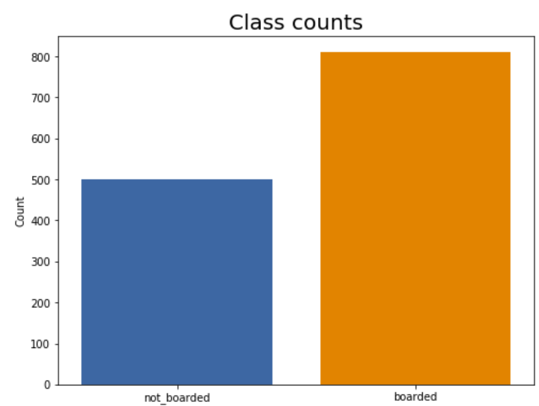


Fig. 2 Class counts of railway data

4.1 Observations

4.1.1 Perceptron Model

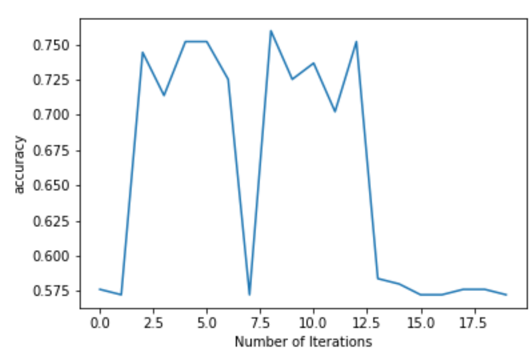


Fig. 3 Variation of Accuracy on Y-axis Vs Number of iterations*500 + 100 on X-axis

Perceptron model gave best accuracy around 4100 iterations with non-standardized data.
Performance on data-
Training set accuracy: 0.7948

Training set f1 score: 0.8456
Test set accuracy: 0.7595
Test set f1 score: 0.8141

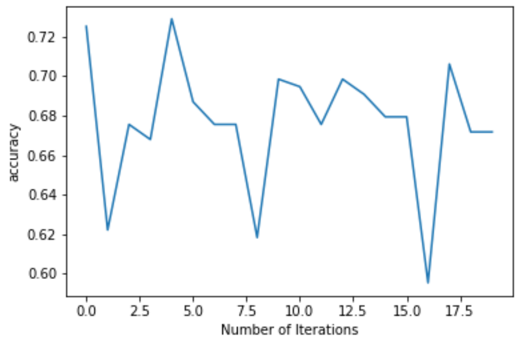


Fig. 4 Accuracy Vs Number of iterations*500 + 100 on scaled features

Best accuracy on non scaled dataset is more than scaled which is very interesting.
Scaled dataset max accuracy around 2100 iterations:0.7290

4.1.2 FLDA Model

Fisher Linear Discriminant Analysis failed miserably on the binary featured dataset with Test set accuracy: 0.4427 Test set f1 score: 0.0519. But the performance was reasonable on scaled dataset.
Performance on scaled data-
Training set accuracy: 0.7299
Training set f1 score: 0.7631
Test set accuracy: 0.7748
Test set f1 score: 0.7915

4.1.3 Logistic Regression(Sigmoid wrapping)

This problem is binary classification, so the defined approach is using sigmoid as the wrapping function. The results using the softmax approach are also listed. On binary features, the sigmoid regression gave poor performance.

Training set accuracy: 0.629
Training set f1 score: 0.772
Test set accuracy: 0.572
Test set f1 score: 0.728
On scaled data:
Training set accuracy: 0.7929
Training set f1 score: 0.8403
Test set accuracy: 0.7862
Test set f1 score: 0.8271

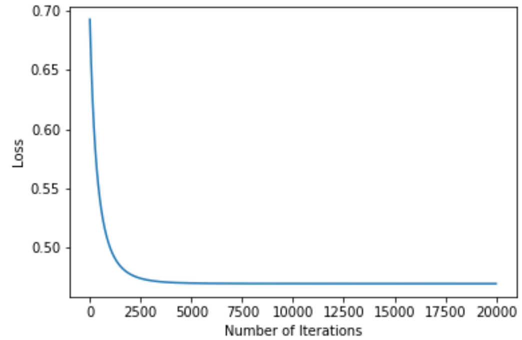


Fig. 5
Loss Vs Number of iterations; alpha=0.005; lambda=0.5

4.1.4 SVM Model

SVM model was tested on the data using different kernels. Only the linear and the Guassian kernel could function on the non-standardized dataset. The results listed are obtained from the standardized dataset.

SVM:

Kernel	Accuracy	F1-score
Linear	0.7595	0.8012
Gaussian	0.7633	0.8165
Polynomial	0.7633	0.8176
sigmoid	0.6793	0.7358

4.2 Conclusion

The discrete and categorical features of this dataset give rise to interesting behaviours. Standardizing the data to zero mean and unit variance improved the performance of FLDA, Logistic regression and SVM models significantly whereas the Perceptron model performed better on binary featured data(Test set accuracy:0.75). All the models performed in the range 0.70-0.80. The best model was obtained using Logistic regression on the scaled dataset achieving accuracy around 0.79 compared to SVM's 0.7633. Not many models give satisfactory results on discrete features except perceptron and SVM. The Perceptron algo does not converge. It fluctuates in regular intervals showing the non-separable nature of the data as can be seen in Fig. 3 and Fig. 4.

5 FASHION MNIST DATA

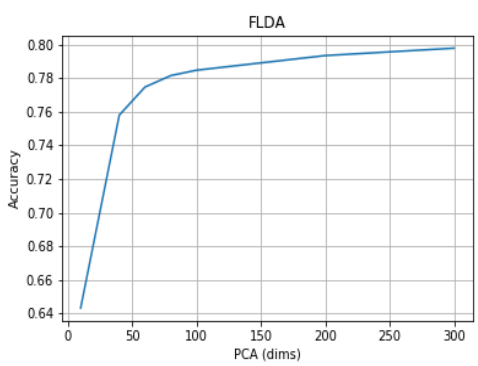
This is a multi-class classification problem where the feature vectors are sparse. Constrained by the memory and computational power, the number of experiments for this data set were reduced. Standardizing the data set with each feature having zero mean and unit variance resulted in unsatisfactory results. Therefore, another normalization technique was used where each example was scaled to have unit norm. Multi class discrimnative models: Perceptron, Logistic, FLDA , SVM were implemented. Here, assume number of components after PCA to be 80 unless stated.

5.1 Observations

5.1.1 Learning Algorithms

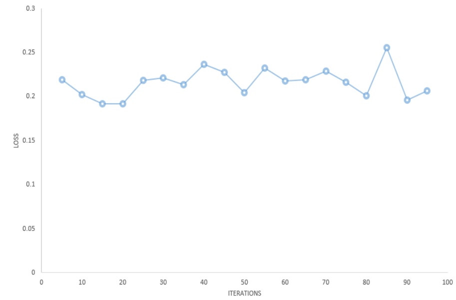
FLDA

Dim.	Acc.	Macro Avg.	Micro Avg.
20	0.713	0.7223	0.713
40	0.7568	0.7623	0.7568
60	0.7747	0.7755	0.7747
80	0.7815	0.7824	0.7815
100	0.784	0.7852	0.7847
200	0.7934	0.7939	0.7934
300	0.7978	0.7972	0.7978



Perceptron

Dims	Acc.	Macro Avg.	Micro Avg.
20	0.5967	0.5848	0.5967
40	0.6971	0.6962	0.6971
60	0.7224	0.7158	0.7224
80	0.7173	0.7098	0.7173
100	0.7219	0.7167	0.7219
200	0.7263	0.7171	0.7263
400	0.6998	0.7099	0.6998
784	0.7981	0.8101	0.7981

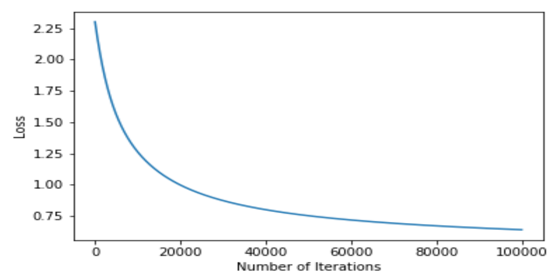


(fig:F2)

SVM:

Kernel	Accuracy	Macro Avg.
Linear	0.8394	0.842
rbf	0.7867	0.7855
Polynomial (degree= 2)	0.6201	0.7542
Polynomial (Degree= 4)	0.31	0.77

Logistic Regression:
Learning rate = 0.012, Iterations =100,000
lambda = 0.0001 Accuracy = 0.778



5.2 Conclusion

The perceptron algorithm could not converge. Error curve on training set (fig:F2) shows the non-monotonic behavior of model indicating that the data is not linearly separable. The best accuracy 0.839 was obtained in case of SVM with linear kernel. Logistic regressions seems to work best when we have regularisation parameter lambda close to zero. The loss curve seems to saturate after 100k iterations. FLDA best accuracy was found to be 0.79 when we retained large number of components (300) after PCA.

5.3 References

[1] Fashion MNIST Reader. Retrieved from https://github.com/zalandoresearch/fashion-mnist/blob/master/utils/mnist_reader.py
[2] Metric for multiclass classification. Retrieved from https://stats.stackexchange.com /questions/51296/