

Assignment 3

Kabir Uberoi - 2022MT61202

November 10, 2024

1 Hard Margin Linear SVM

1.1 Number of Support Vectors

The number of support vectors used in the trained Hard Margin Linear SVM model is 10004.

Number of Support Vectors: 10004

1.2 Top 6 Support Vectors

The top 6 support vectors for the Hard Margin Linear SVM are shown below. These support vectors are the most influential in defining the decision boundary of the model.

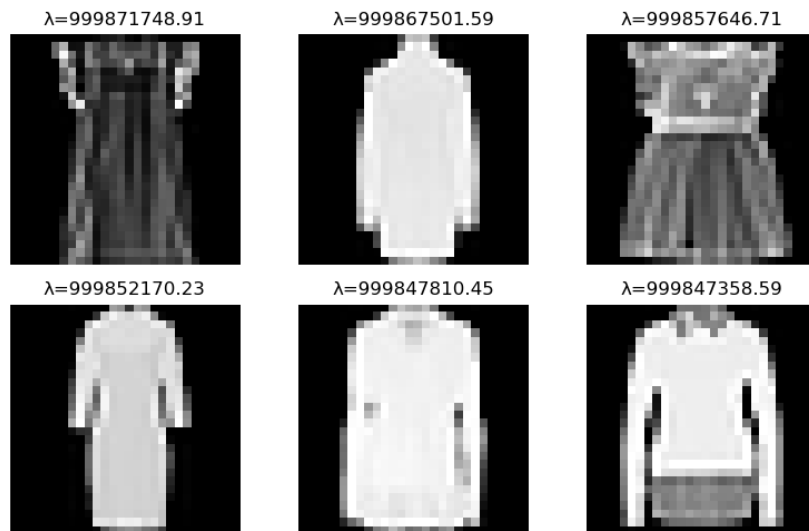


Figure 1: Top 6 Support Vectors for Hard Margin Linear SVM

1.3 Performance Metrics

1.3.1 Training Accuracy and F1-Score

For the Hard Margin Linear SVM, the following training performance metrics are reported:

Training Accuracy: 0.8139

Training F1-Score: 0.7213

1.3.2 Validation Accuracy and F1-Score

The performance on the validation set for the Hard Margin Linear SVM is as follows:

Validation Accuracy: 0.7483

Validation F1-Score: 0.70

1.4 Misclassified Instances

1.4.1 Training Misclassified Instances

Below are images showing four misclassified instances from the training set:

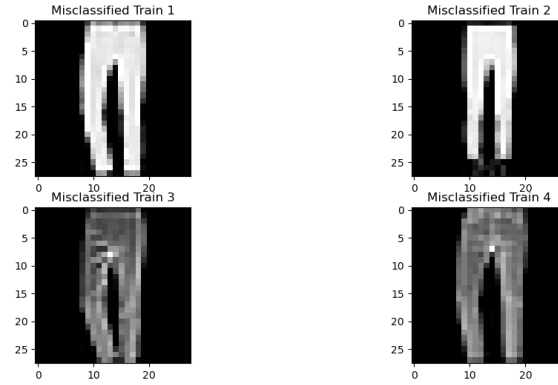


Figure 2: Misclassified Instances (Train) for Hard Margin Linear SVM

1.4.2 Validation Misclassified Instances

Below are images showing four misclassified instances from the validation set:

2 Soft Margin Linear SVM

2.1 Number of Support Vectors

The optimal C was 0.01. The number of support vectors used in the trained Soft Margin Linear SVM model is 10004.

Number of Support Vectors: 10004

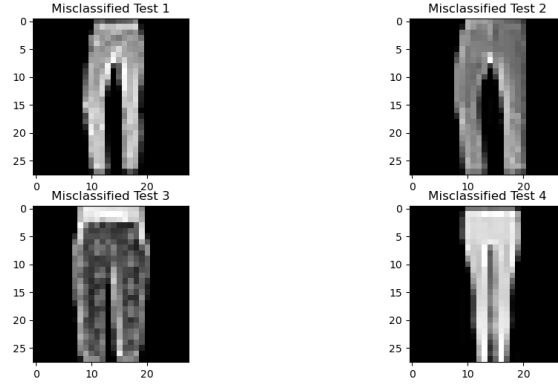


Figure 3: Misclassified Instances (Validation) for Hard Margin Linear SVM

2.2 Top 6 Support Vectors

The top 6 support vectors for the Soft Margin Linear SVM are shown below. These support vectors are the most influential in defining the decision boundary of the model.

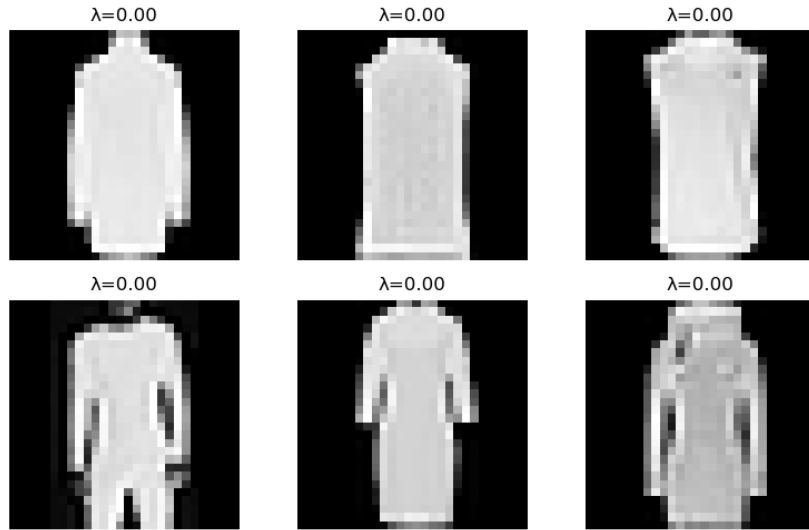


Figure 4: Top 6 Support Vectors for Soft Margin Linear SVM

2.3 Performance Metrics

2.3.1 Training Accuracy and F1-Score

For the Soft Margin Linear SVM, the following training performance metrics are reported:

Training Accuracy: 0.8818

Training F1-Score: 0.7838

2.3.2 Validation Accuracy and F1-Score

The performance on the validation set for the Soft Margin Linear SVM is as follows:

Validation Accuracy: 0.8769

Validation F1-Score: 0.8278

2.4 Misclassified Instances

2.4.1 Training Misclassified Instances

Below are images showing misclassified instances from the training set:

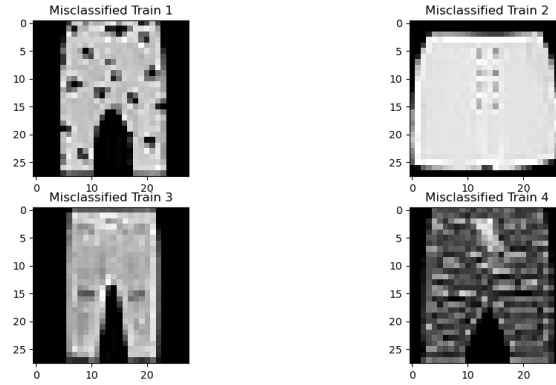


Figure 5: Misclassified Instances (Train) for Soft Margin Linear SVM

2.4.2 Validation Misclassified Instances

Below are images showing misclassified instances from the validation set:

3 Soft Margin RBF SVM

3.1 Number of Support Vectors

The optimal C was 10. The number of support vectors used in the trained Soft Margin RBF SVM model is 10004.

Number of Support Vectors: 10004

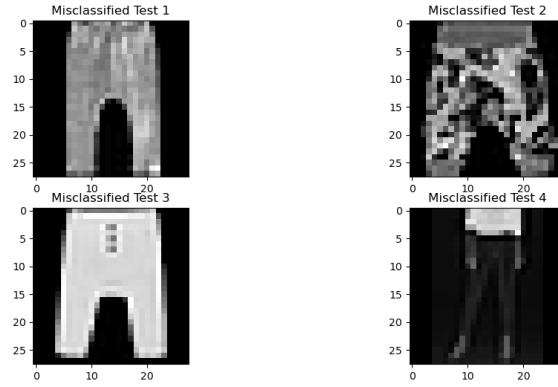


Figure 6: Misclassified Instances (Validation) for Soft Margin Linear SVM

3.2 Top 6 Support Vectors

The top 6 support vectors for the Soft Margin RBF SVM are shown below. These support vectors are the most influential in defining the decision boundary of the model.

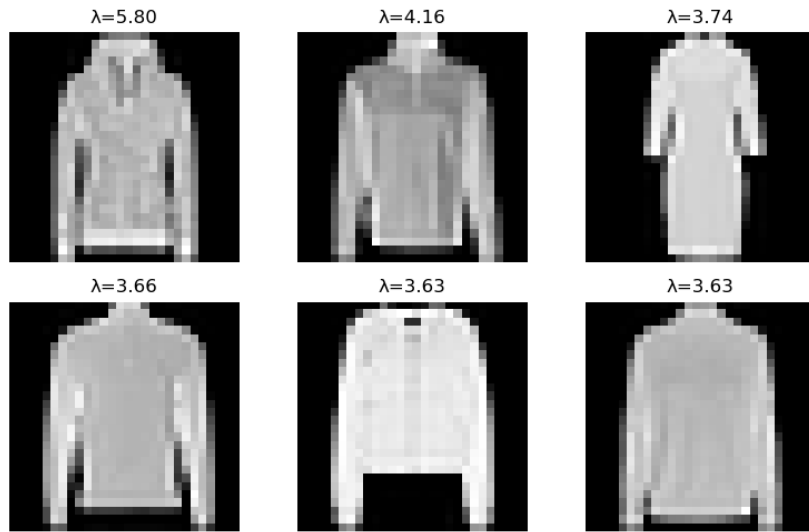


Figure 7: Top 6 Support Vectors for Soft Margin RBF SVM

3.3 Performance Metrics

3.3.1 Training Accuracy and F1-Score

For the Soft Margin RBF SVM, the following training performance metrics are reported:

Training Accuracy: 1.0

Training F1-Score: 1.0

3.3.2 Validation Accuracy and F1-Score

The performance on the validation set for the Soft Margin RBF SVM is as follows:

Validation Accuracy: 0.9344

Validation F1-Score: 0.8919

3.4 Misclassified Instances

3.4.1 Training Misclassified Instances

As the F1 is 1, there are no misclassifications

3.4.2 Validation Misclassified Instances

Below are images showing misclassified instances from the validation set:

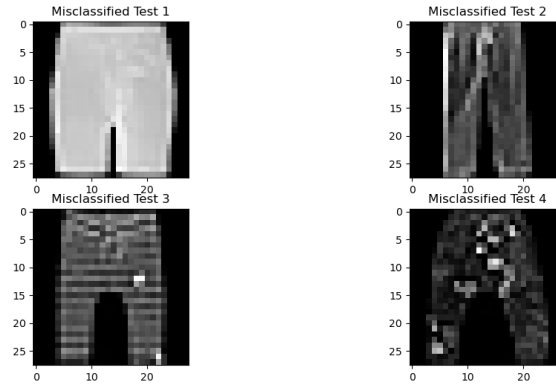


Figure 8: Misclassified Instances (Validation) for Soft Margin RBF SVM

4 Hard Margin RBF SVM

4.1 Number of Support Vectors

The number of support vectors used in the trained Hard Margin RBF SVM model is 10004.

Number of Support Vectors: 10004

4.2 Top 6 Support Vectors

The top 6 support vectors for the Hard Margin RBF SVM are shown below. These support vectors are the most influential in defining the decision boundary of the model.

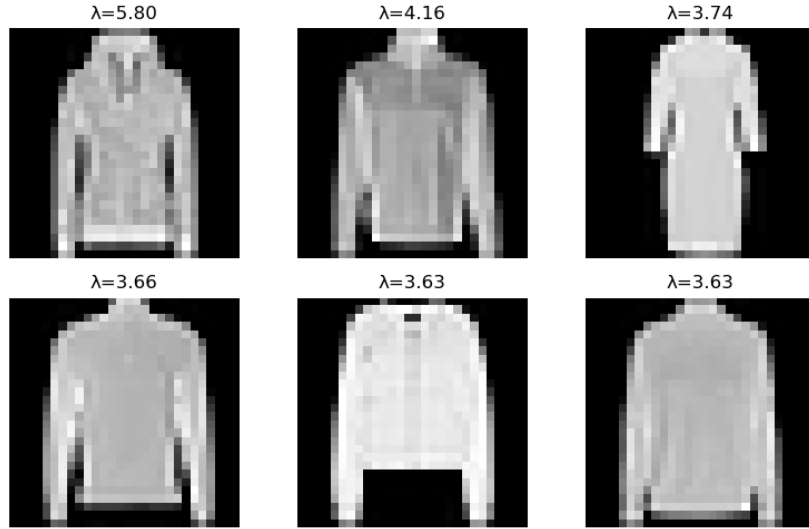


Figure 9: Top 6 Support Vectors for Hard Margin RBF SVM

4.3 Performance Metrics

4.3.1 Training Accuracy and F1-Score

For the Hard Margin RBF SVM, the following training performance metrics are reported:

Training Accuracy: 1.0

Training F1-Score: 1.0

4.3.2 Validation Accuracy and F1-Score

The performance on the validation set for the Hard Margin RBF SVM is as follows:

Validation Accuracy: 0.9344

Validation F1-Score: 0.8919

4.4 Misclassified Instances

As the F1 is 1, there are no misclassifications

4.4.1 Validation Misclassified Instances

Below are images showing misclassified instances from the validation set:

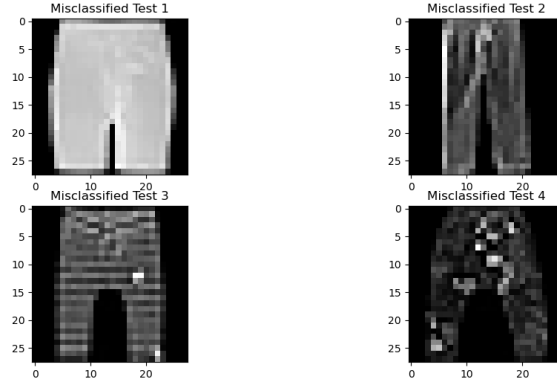


Figure 10: Misclassified Instances (Validation) for Hard Margin RBF SVM

5 Random Forest

The hyper-parameters I used are Depth = 50, Number of Principle Components = 58 (This was a time and accuracy trade off), number of trees = 20 (Time Constraint) and Minimum Splitting size of node = 5. These were found using Grid Search which I tweaked manually to get a reasonably good model.

The Model is underfitting but higher hyper-parameters are infeasible due to the time constraints.

5.1 Performance Metrics

5.1.1 Training Accuracy and F1-Score

For the Random Forest model, the following training performance metrics are reported:

Training Accuracy: 0.8713

Training F1-Score: 0.7772

5.1.2 Test Accuracy and F1-Score

The performance on the test set for the Random Forest model is as follows:

Test Accuracy: 0.8202

Test F1-Score: 0.7491

5.2 Misclassified Instances

5.2.1 Training Misclassified Instances

Below are images showing misclassified instances from the training set:

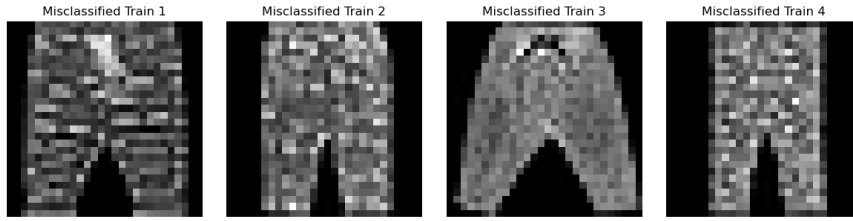


Figure 11: Misclassified Instances (Train) for Random Forest

5.2.2 Test Misclassified Instances

Below are images showing misclassified instances from the test set:

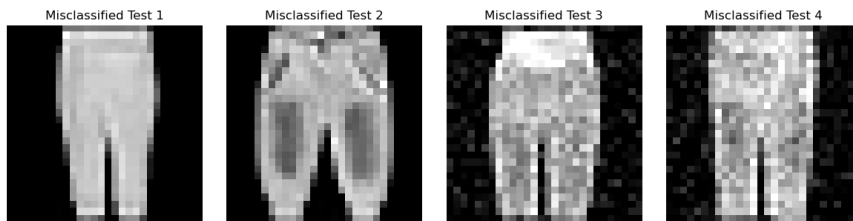


Figure 12: Misclassified Instances (Test) for Random Forest

6 AdaBoost Classifier

The hyper-parameters I used are Depth = 1, Number of Principle Components = 114 (As 114 gives 95% variance.) and number of trees = 200 (Time Constraint)

6.1 Performance Metrics

6.1.1 Training Accuracy and F1-Score

For the AdaBoost model, the following training performance metrics are reported:

Training Accuracy: 0.9126

Training F1-Score: 0.8200

6.1.2 Test Accuracy and F1-Score

The performance on the test set for the AdaBoost model is as follows:

Test Accuracy: 0.9006

Test F1-Score: 0.8375

6.2 Misclassified Instances

6.2.1 Training Misclassified Instances

Below are images showing misclassified instances from the training set:

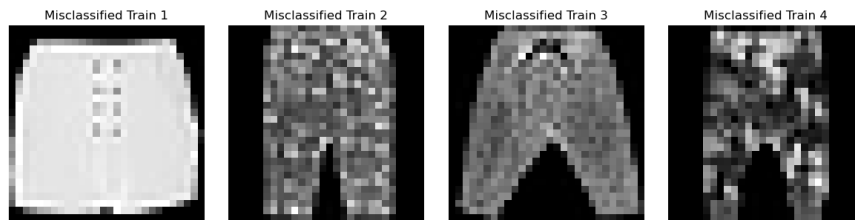


Figure 13: Misclassified Instances (Train) for AdaBoost

6.2.2 Test Misclassified Instances

Below are images showing misclassified instances from the test set:

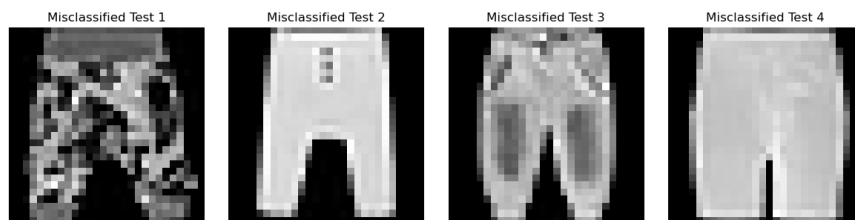


Figure 14: Misclassified Instances (Test) for AdaBoost