

Report: Exploring Linear Discriminant Analysis (LDA) and Naive Bayes Classifier

Atharva Date - B22AI045

1. Introduction

This report explores the application of Linear Discriminant Analysis (LDA) and Naive Bayes theorem in analyzing and classifying data.

2. Linear Discriminant Analysis (LDA)

2.1 Data Description

The dataset comprises 2000 entries, evenly split between values 1 and 0.

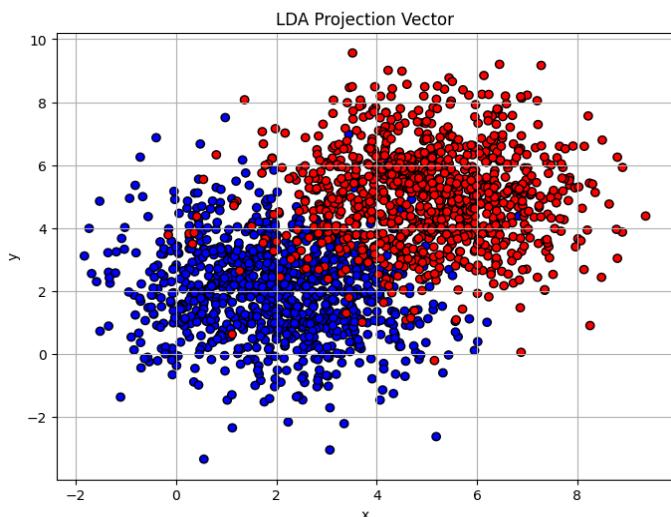
Variance and Mean of X and Y coordinates of Data with Value = 0:

- Variance: (2.1925, 2.2735)
- Mean: (2.0492, 2.0277)

Variance and Mean of X and Y coordinates of Data with Value = 1:

- Variance: (2.2285, 2.2829)
- Mean: (5.0348, 5.0520)

The 1s are located predominantly in the top-right part of the graph,

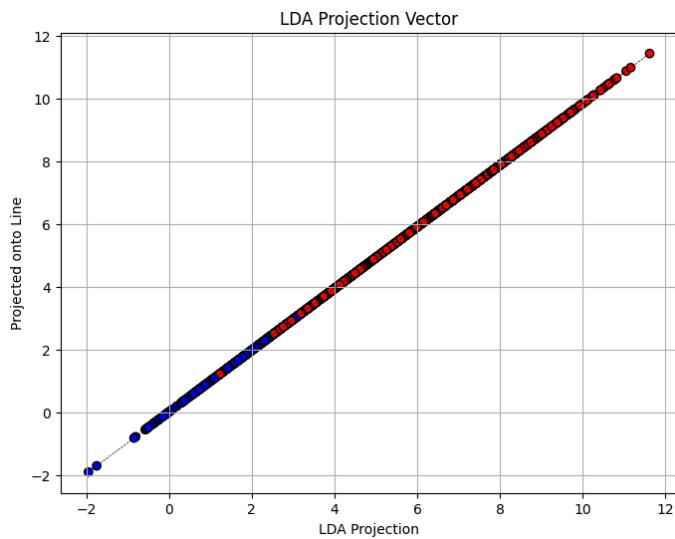


while the 0s are concentrated in the bottom-left part. 1s have higher variance, indicating they are more spread out.

2.2 Plotting LDA Projection Vector

Variance and Mean of Projected Data:

- Value = 0: Variance = 2.0237, Mean = 2.8829
- Value = 1: Variance = 2.2265, Mean = 7.1320



Analysis:

- The variance of data points labeled as 0 remains similar between the original and projected data. However, the variance for data points labeled as 1 remains unchanged.
- The mean of data points labeled as 0 shifts slightly from approximately (2.05, 2.03) in the original data to approximately (2.88) in the projected data. Similarly, the mean of data points labeled as 1 shifts from approximately (5.03, 5.05) in the original data to approximately (7.13) in the projected data.
- These shifts in mean values suggest that the projection has caused a displacement of the data points along the direction of the projection vector.

Insights:

- The variance analysis indicates that the dispersion of data points remains relatively consistent after projection, particularly for data points labeled as 1.
- However, the shift in mean values suggests that the projection has influenced the central tendency of the data points, changing their distribution.
- These changes in variance and mean values contribute to the observed differences in the spread and coverage of data points between the original

and projected spaces, as evidenced by the analysis of the confusion matrix and prediction accuracy.

2.3 Comparison with 1-NN Classifier

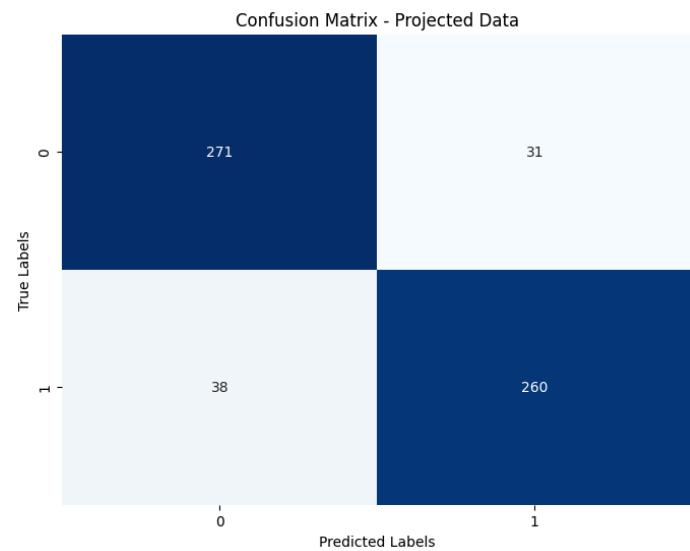
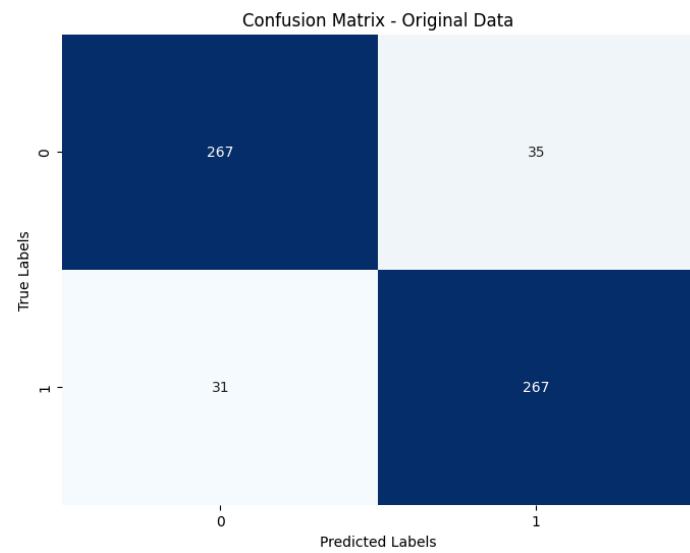
Spread of Data: The region covered by samples labeled as 0 in the projected data appears to be smaller compared to the original data. This suggests that the projection has concentrated the samples labeled as 0 into a smaller area of the feature space.

Confusion Matrix Analysis: When comparing the confusion matrices between the original and projected data, we observe that:

The number of true 1 values predicted as 0 (false negatives) and true 0 values predicted as 0 (true negatives) is higher in the projected data compared to the original data.

Conversely, the number of true 0 values predicted as 1 (false positives) and true 1 values predicted as 1 (true positives) is lower in the projected data compared to the original data.

Accuracy:



- **Original Data: 0.89**
- **Projected Data: 0.885**

Prediction Accuracy: The prediction accuracy of the model on the projected data is lower than that on the original data. This indicates that the projection has potentially caused some loss of information or separation between the classes, leading to reduced predictive performance.

Overall, the projection appears to have affected the distribution and separability of the data, resulting in a different performance of the classification model. The reduction in spread and coverage of samples in the projected space, along with the changes in the confusion matrix and prediction accuracy, suggest that the projection may not have fully captured the discriminative information present in the original data.

3. Naive Bayes Classifier

3.1 Task Description

It involves implementing a Naive Bayes classifier and calculating prior, likelihood, and posterior probabilities, prediction and changes observed due to laplace smoothing.

3.2 Implementation Details

The classifier is implemented by calculating prior probabilities, likelihood probabilities for each feature given the class, and posterior probabilities. Laplace smoothing is incorporated to handle zero probabilities.

3.3 Prior Probabilities

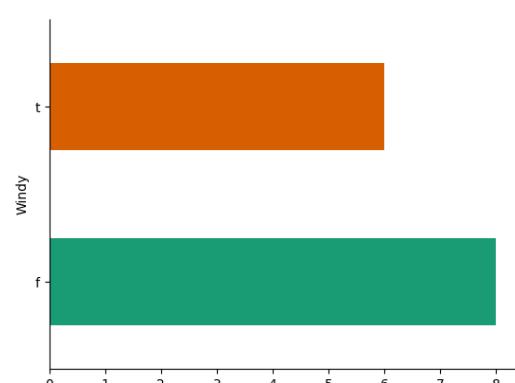
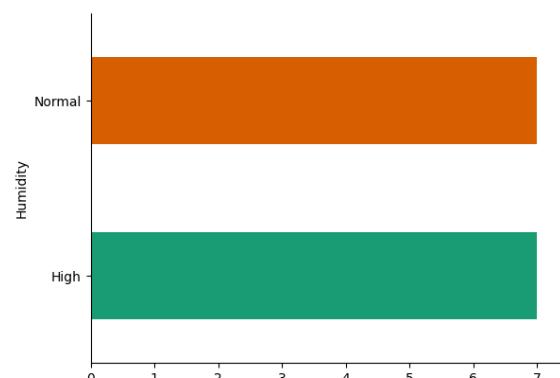
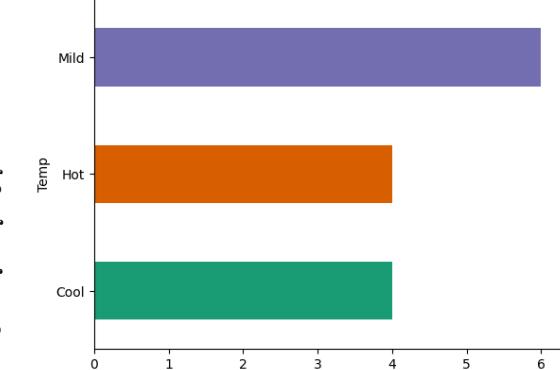
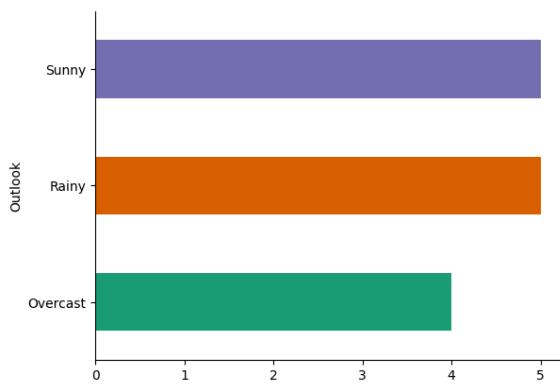
- 'no': 0.3333
- 'yes': 0.6667

3.4 Likelihood Probabilities

Class: no

- Feature 0: Rainy (0.5), Sunny (0.5)
- Feature 1: Cool (0.25), Hot (0.25), Mild (0.5)
- Feature 2: High (0.75), Normal (0.25)
- Feature 3: f (0.5), t (0.5)

Class: yes



- Feature 0: Overcast (0.5), Rainy (0.25), Sunny (0.25)
- Feature 1: Cool (0.375), Hot (0.25), Mild (0.375)
- Feature 2: High (0.25), Normal (0.75)
- Feature 3: f (0.625), t (0.375)

3.5 Posterior Probabilities

- $\{ \{ \text{'no'} : 0.0078125, \text{'yes'} : 0.009765625 \} \}$
- $\{ \text{'no'} : 0.0078125, \text{'yes'} : 0.009765625 \}$

3.6 Predictions

- ‘yes’
- ‘yes’

3.7 Using Laplace Smoothing

Incorporating Laplace smoothing ensures that even feature-value combinations with zero probabilities receive a non-zero probability estimate, thus preventing issues during classification. Specifically, when there is an 'Overcast' condition but no playing event, Laplace smoothing helps assign a likelihood probability of 0.125.

Class: no

- Feature 0: Rainy (0.375), Sunny (0.375), Overcast (0.125)
- Feature 1: Cool (0.25), Hot (0.375), Mild (0.25)
- Feature 2: High (0.5), Normal (0.25)
- Feature 3: f (0.25), t (0.5)

Class: yes

- Feature 0: Overcast (0.4167), Rainy (0.25), Sunny (0.25)
- Feature 1: Cool (0.3333), Hot (0.25), Mild (0.3333)
- Feature 2: High (0.25), Normal (0.5833)
- Feature 3: f (0.5), t (0.3333)

There is not any difference between the test results as the test data does not contain the case where the probability result by the bayes theorem is such that the end result in spite of it being Overcast is not playing.

But it was also observed that on trying to predict all values of the dataset we get that both models have yielded the same results. Two instances were misclassified as 'yes' when they should have been classified as 'no'. This suggests that the classifier might be making errors in predicting the negative class ('no') due to insufficient representation of 'no' values in the dataset.

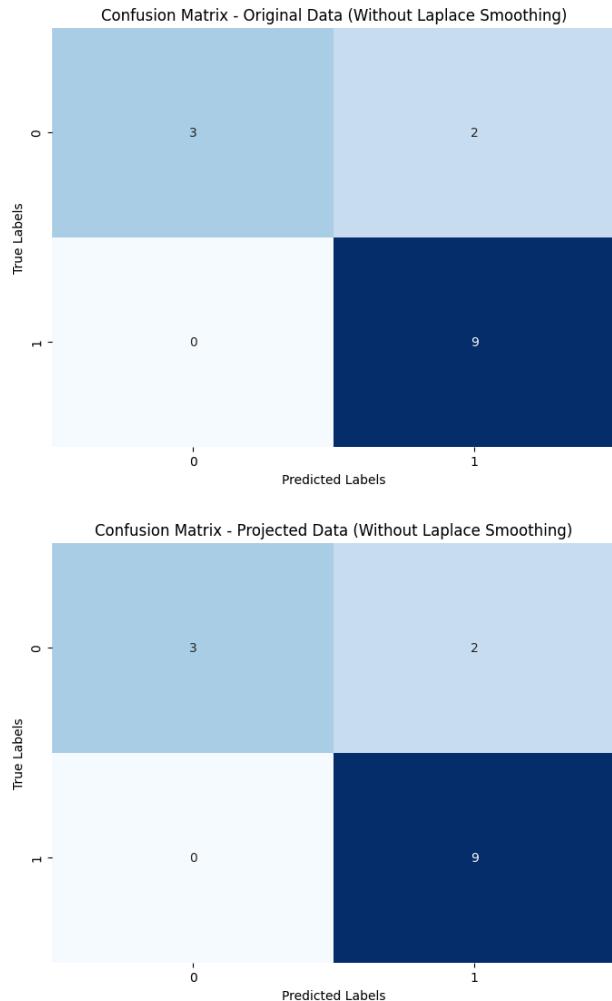
Difference in Posterior Probabilities

- Post: `[{'no': 0.00390625, 'yes': 0.00694444444444444}]`
- `{'no': 0.00390625, 'yes': 0.00694444444444444}]`

As we can observe the probability of no has increased in both the cases. We can sense that the decision making of the model has become more balanced as compared to the previous case.

3.8 Analysis

In summary, the Naive Bayes classifier does a good job of figuring out how features relate to different categories. Adding Laplace smoothing makes the classifier more reliable when dealing with rare or unseen situations. The likelihood probabilities help us understand how likely certain features are for each category, which helps



us make better predictions. Overall, the Naive Bayes classifier is a useful tool for making accurate predictions in various situations.