

Q.2 Using Naïve Bayes

Abalone - raw dataset - Multinomial naive bayes

If we apply Standardisation to the Abalone dataset, values become negative and that is not acceptable as a values to Naive Bayes classifiers. Hence, we need to use MinMaxScaler (Normalization) to scale down values only within 0 and 1. However, this will decrease the accuracy of the model.

The accuracy of a model on the Raw abalone dataset has significantly reduced from 26% to 16.5% with Naive Bayes compared to KNN using 10 neighbors measured in the previous assignment. While it's likely that neither algorithm is adequate for predicting the abalone age, the KNN model is more accurate so far

Wine Raw Dataset using multinomial Naïve Bayes: Mean Accuracy

On Wine dataset, we got mean accuracy of 46.159%

KNN Algorithm has worked slightly better on the Wine (Raw) dataset compared to Multinomial Naive Bayes as the accuracy has gone down from 46.15% to an average of 41.5% accross 5-folds. A combination of Standardisation and then KNN has no significant effect on the accuracy improvement.

Using Raw complement NB on Abalone dataset, we got the accuracy of 17.5%

Using Raw complement NB on Wine dataset, we got the accuracy of 38.971%

Following are the tables for accuracies:

Test accuracy of Raw abalone

Cross-validation accuracy of Raw abalone using Multinomial Naive Bayes classifier: 16.37%

Cross-validation accuracy of Raw abalone using Complement Naive Bayes classifier: 18.14%

Test accuracy of Raw abalone using Multinomial Naive Bayes classifier: 16.99%

Test accuracy of Raw abalone using Complement Naive Bayes classifier: 19.14%

Test accuracy of PCA processed abalone

Cross-validation accuracy of processed abalone using Multinomial Naive Bayes classifier with PCA: 16.37%

Cross-validation accuracy of processed abalone using Complement Naive Bayes classifier with PCA: 18.26%

Test accuracy of processed abalone using Multinomial Naive Bayes classifier with PCA: 16.99%

Test accuracy of processed abalone using Complement Naive Bayes classifier with PCA: 17.22%

Test accuracy of LDA processed abalone

Cross-validation accuracy of processed abalone using Multinomial Naive Bayes classifier with LDA: 16.37%

Cross-validation accuracy of processed abalone using Complement Naive Bayes classifier with LDA: 23.97%

Test accuracy of processed abalone using Multinomial Naive Bayes classifier LDA: 16.99%

Test accuracy of processed abalone using Complement Naive Bayes classifier LDA: 21.53%

Wine Dataset implementation

Test Accuracy of raw wine

Cross-validation accuracy of raw wine using Multinomial Naive Bayes classifier: 43.54%

Cross-validation accuracy of raw wine using Complement Naive Bayes classifier: 47.51%

Test accuracy of raw wine using Multinomial Naive Bayes classifier: 44.62%

Test accuracy of raw wine using Complement Naive Bayes classifier: 48.54%

Test accuracy of PCA processed wine

Cross-validation accuracy of processed wine using Multinomial Naive Bayes classifier with PCA: 43.43%

Cross-validation accuracy of processed wine using Complement Naive Bayes classifier with PCA: 45.41%

Test accuracy of processed wine using Multinomial Naive Bayes classifier with PCA: 44.54%

Test accuracy of processed wine using Complement Naive Bayes classifier with PCA: 46.92%

Test accuracy of LDA processed wine

Cross-validation accuracy of processed wine using Multinomial Naive Bayes classifier with LDA: 43.43%

Cross-validation accuracy of processed wine using Complement Naive Bayes classifier with LDA: 0.54%

Test accuracy of processed wine using Multinomial Naive Bayes classifier with LDA: 44.54%

Test accuracy of processed wine using Complement Naive Bayes classifier with LDA: 0.15%

Conclusion for Abalone Dataset

The above accuracies summarize the test accuracy of two algorithms (Multinomial Naive Bayes and Complement Naive Bayes) on three versions of the abalone dataset: raw, PCA-preprocessed (with 3 principal components), and LDA-preprocessed (with 3 linear discriminants).

For **Multinomial Naive Bayes**, the test accuracy remains the same (16.99%) across all three versions of the dataset.

For **Complement Naive Bayes**, the test accuracy is highest on the LDA-preprocessed dataset (21.53%), followed by the raw dataset (19.14%), and lowest on the PCA-preprocessed dataset (17.22%). This suggests that LDA pre-processing is more effective for improving the performance of Complement Naive Bayes on the abalone dataset compared to PCA pre-processing.

Conclusion for Wine Dataset

For the **Multinomial Naive Bayes** algorithm, there is not much difference in performance between the raw wine dataset and the dataset preprocessed with PCA or LDA. The accuracy remains around 44-45% for all three settings.

For the **Complement Naive Bayes** algorithm, the performance is significantly better on the raw wine dataset compared to the preprocessed datasets. The accuracy is around 48.5% for the raw dataset, but drops to around 47% for the dataset preprocessed with PCA, and drops even further to 0.15% for the dataset preprocessed with LDA.