DD2421: MACHINE LEARNING LAB3

Author
Adel Abdelsamed

Contents

1 Solutions

Solution Assignment 1: The maximum likelihood estimate of a given dataset is computed and the corresponding 95% confidence intervals are plotted in Figure1.

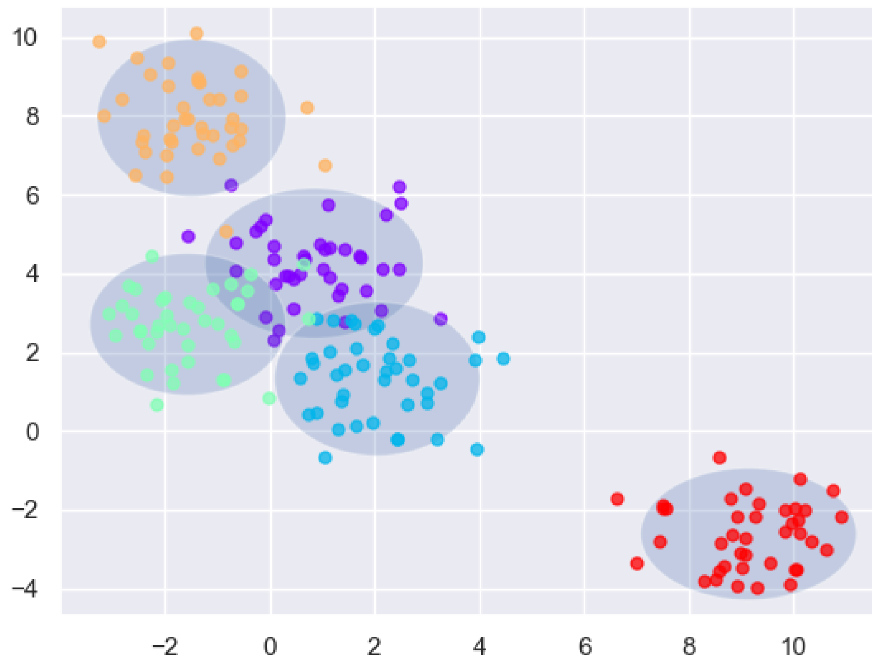

Figure 1: Gaussian distributions for the Maximum Likelihood estimate.

> **Q1:**: When can a feature independence assumption be reasonable and when not?

Q1 Solution: The feature independence assumption in a Naive Bayes classifier is a simplifying assumption that is made for computational ease. It assumes that the features used to classify data are conditionally independent given the class label. In other words, it assumes that the presence or absence of one feature does not affect the presence or absence of any other feature.

Hence, the use of the Naive Bayes classifier is data dependent and is reasonable whenever the correlations of the data is not significant.

> **Q2:** How does the decision boundary look for the Iris dataset? How could one improve the classification results for this scenario by changing classifier or, alternatively, manipulating the data?

Q2 Solution: The plot is shown in Figure 2. The decision boundary for class 0 is captured well. However, for class 1 and 2 the decision boundary is not accurate. Since both classes have overlapping data points, this decision boundary is clearly more complex to learn. The inaccurate decision boundary is a result of the simple Naive Bayes assumption. The classifier could be improved if feature independence is considered. Another alternative would be to use SVM with a nonlinear kernel to account for the overlap between the data sets. The result of Naive Bayes could be improved by transforming the dataset into a higher dimensional space, which would increase the computational load.

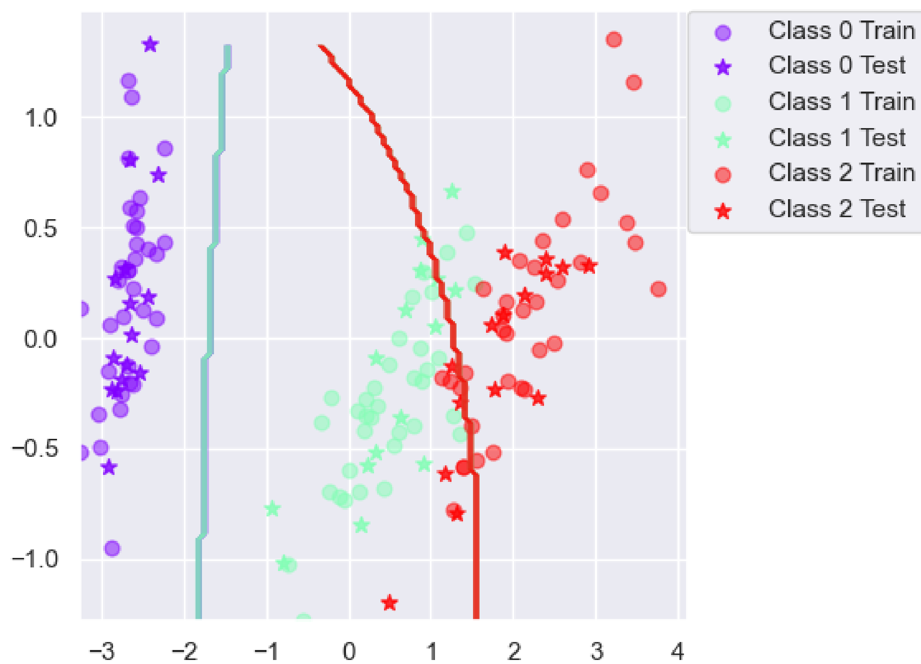

Figure 2: Decision Boundary for the Iris dataset using Naive Bayes Classifier: Final mean classification accuracy 89 with standard deviation 4.16.

The Naive Bayes Classifier was utilized and applied within the Adaboost algorithm. The Naive Bayes classifier was found to perfrom poorly. Thus, it is combined with an ensemble-based boosting, where multiple weak learners are utilized to create a stronger, more accurate classifier.

Q3: Is there any improvement in classification accuracy? Why/why not?

Q3 Solution: The boosted decision boundary has a higher mean classification accuracy and a slightly lower standard deviation (as shown in Figure 3). Since boosting iteratively adjusts the weights of training instances to give more importance to those that are misclassified by the previous weak learners, the overall accuracy is improved (focus on misclassified data). Furthermore, each new weak learner is trained to correct the mistakes of the ensemble so far, which contributes to a more complex decision boundary. In our case the ensemble method reduces the bias resulting from the individual weak learners (Naive Bayes.)

Q4: Plot the decision boundary of the boosted classifier on iris and compare it with that of the basic. What differences do you notice? Is the boundary of the boosted version more complex?

Q4 Solution: The boosted Naive Bayes Classifier generates the decision boundary in Figure 3. The decision boundary between class 0 and 1 is more accurate. The significant improvement, however, lies in the decision boundary between class 1 and 2, where the data is overlapping. The new decision boundary is more complex and manages to obtain a good balance between model complexity and the data noise.

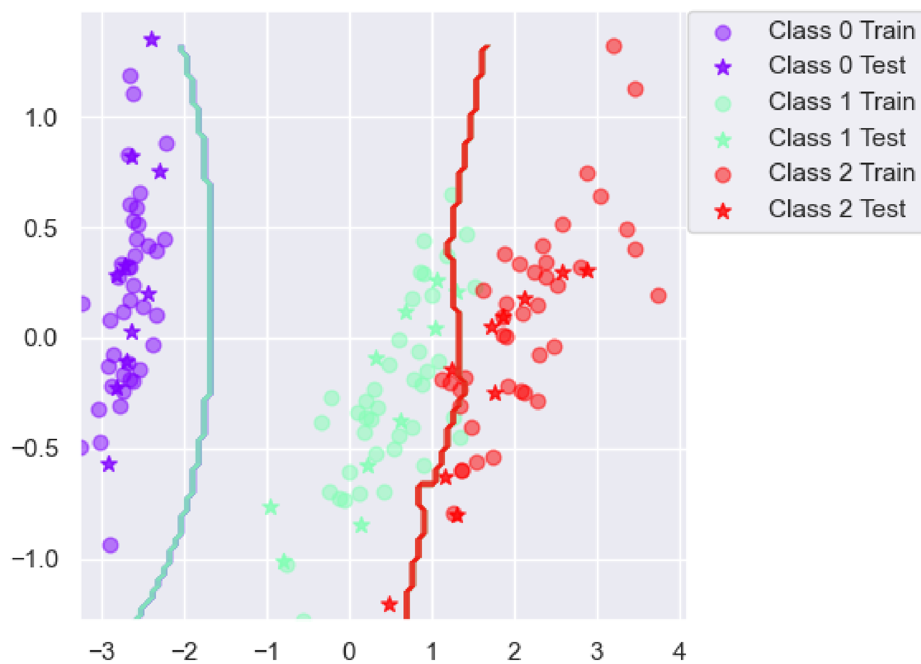

Figure 3: Decision Boundary for the boosted Iris dataset using Naive Bayes Classifier: Final mean classification accuracy 94.2 with standard deviation 4.12.

Q5: Can we make up for not using a more advanced model in the basic classifier (e.g. independent features) by using boosting?

Q5 Solution: Yes, boosting the Naive Bayes Classifier resulted in model with a more complex decision boundary hence clearly lowering the bias of the individual weak learners. This compensates for the feature independence assumption from the Naive Bayes.

Q6: Repeat Q3-Q5 with a decision tree classifier.

The weak learners are now formed using a decision tree classifier. The same observations and conclusions in Q3-Q5 apply here similarly.

Figure 4: Decision Boundary for the Iris dataset using Decision Tree Classifier: Final mean classification accuracy 93 with standard deviation 4.87.

The decision tree classifier is also tested on the vowels data sets. The achieved performance is exhibited below:
Final mean classification accuracy 64.1 with standard deviation 4
Final mean classification accuracy 86.7 with standard deviation 2.7 (Boosted)

The boosted classifier again improves the classification accuracy.

Q7: If you had to pick a classifier, naive Bayes or a decision tree or the boosted versions of these, which one would you pick? Motivate from the following criteria:

1. **Outliers:** Decision Trees tend to overfit in general as seen in Lab1. With proper pruning the variance of the trained tree will decrease making the classifier robust towards noise. Boosting weights classified training samples more and hence might lead to a classifier with high variance. Using Naive Bayes Classifier would hence be the first recommendation.

2. **Irrelevant inputs: part of the feature space is irrelevant** Naive Bayes uses the features to form a likelihood model given the data distributions. The features are treated equally. Decision Trees is a greedy-method, i.e. it chooses the feature which maximizes information gain. With proper pruning the relevant inputs can be kept and the irrelevant inputs will be pruned from the tree. Depending on the performance of the trained tree, boosting can be considered to improve performance.

3. **Predictive power:** Based on the experience from Lab3, the boosted version of either Naive Bayes or Decision Trees classifiers is capable of generating good decision

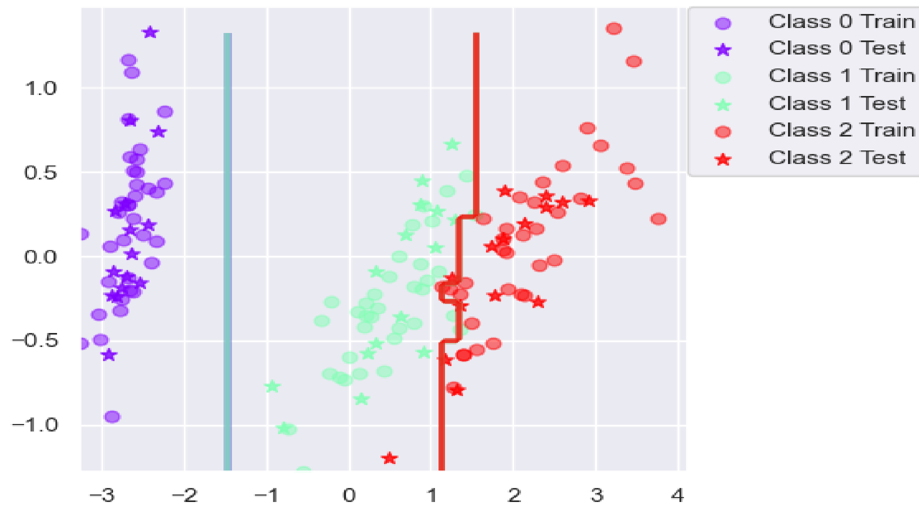

Figure 5: Decision Boundary for the boosted Iris dataset using Decision Tree Classifier: Final mean classification accuracy 94.6 with standard deviation 3.65.

boundaries striking a good balance between complexity of the boundary and robustness to noise. In general, the performance of Naive Bayes depends on the assumption of the underlying data distribution.

4. **Mixed types of data: binary, categorical or continuous features** Both decision trees and the naive Bayes can handle non-continuous features. If the dataset contains mixed data types, proper preprocessing is needed to use Naive Bayes (e.g. Naive Binomial/Categorical Bayes). Decision Trees offer an advantage here as it can deal with mixed data types easily within the same tree.

5. **Scalability: the dimension of the data D is large** In general, using Naive Bayes is computationally efficient and, as seen, the training as well as the inference depends only on some counting operations. However, if the feature space is big and the dataset is large, assuming an independent distribution for each individual feature will decrease the efficiency (curse of dimensionality). In these cases, decision trees are more efficient. In general, using boosting will incur more computational cost. Hence if the dataset is large, ensemble methods such as boosting are not favorable.