

Kieran Broekhoven - 20502763
Nicholas Manna - 20490473
Naveen Nirmalaraj - 20509461

SYDE 372 - Pattern Recognition

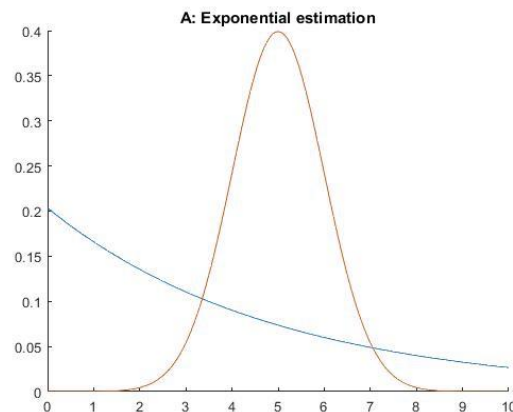
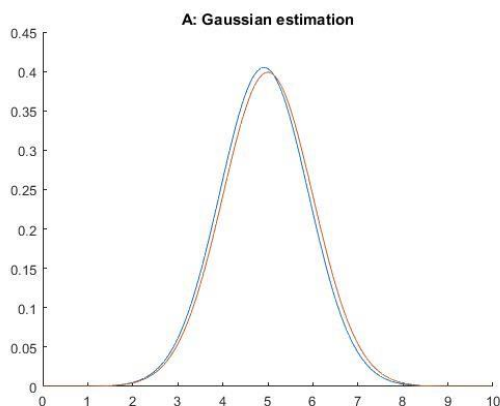
Lab 2

1.0 Introduction

This report assesses various methods of model estimation, first in one dimension. Maximum likelihood parametric estimation is used assuming Gaussian, exponential, and uniform distributions. Non-parametric estimation is also performed using the Parzen method. Afterwards, the two dimension case is analyzed. Parametric estimation is applied assuming a Gaussian distribution. Non-parametric estimation is then used via a Gaussian Parzen window. Lastly, the effects of combining multiple discriminants in a sequential classifier are examined. Aggregating multiple weak classifiers can lead to a more powerful classifier

2.0 Model Estimation 1-D Case

Two data sets are used to test model estimation methods: a Gaussian (A) and an exponential (B) distribution. The following plots for distribution A (Figure 1) and distribution B (Figure 2) were the result of each estimate on each data set.



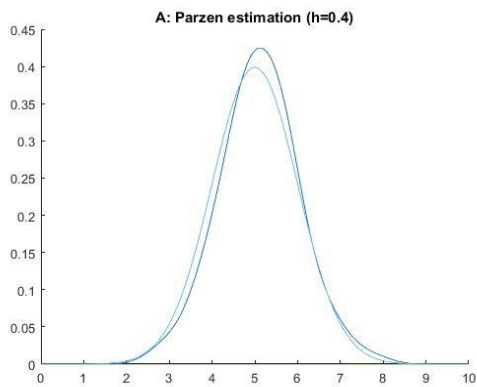
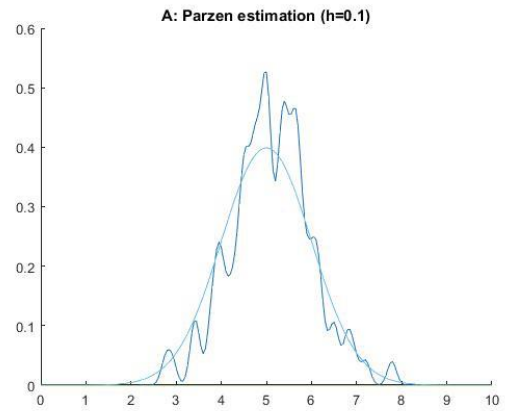
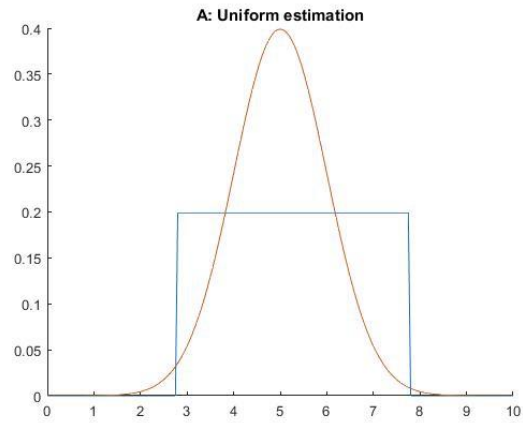
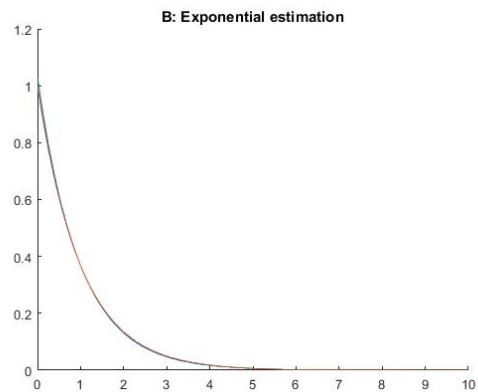
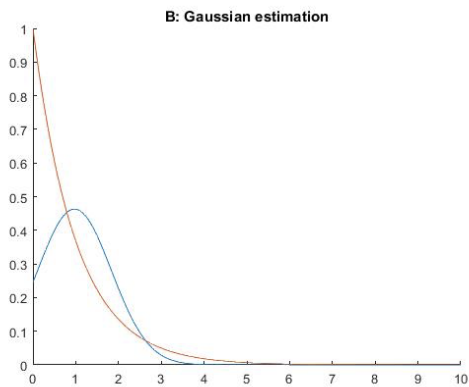


Figure 1: Model estimation on data set A



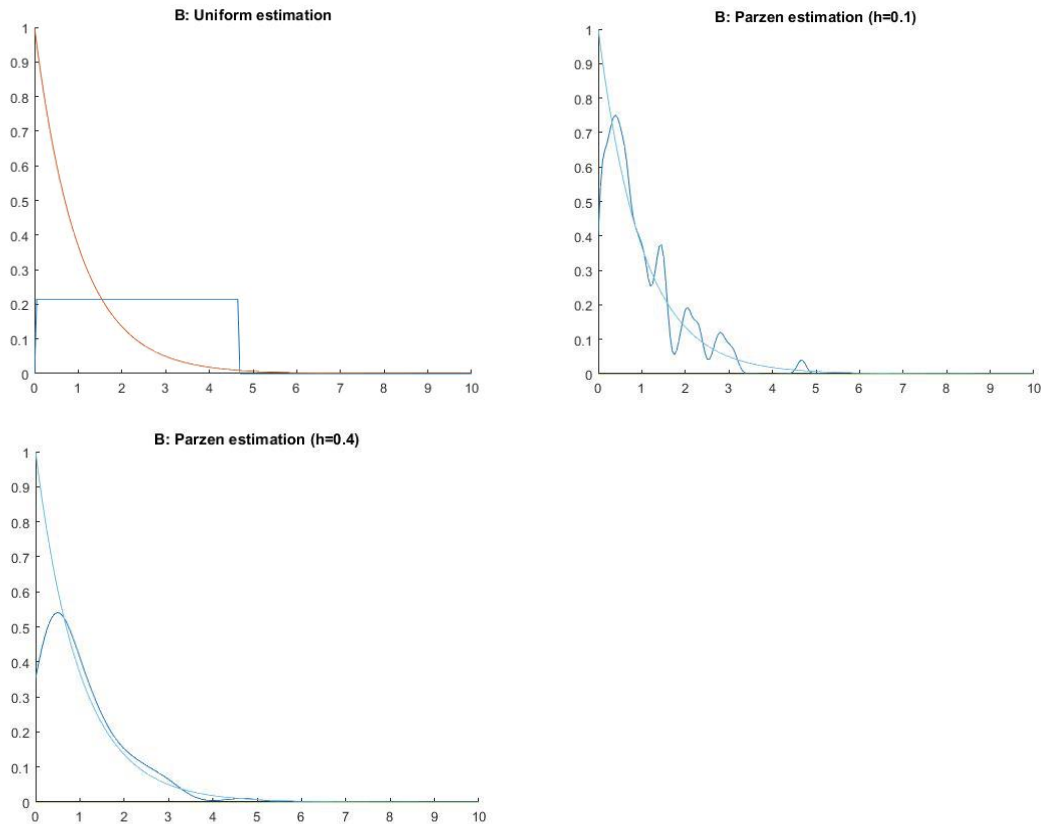


Figure 2: Model estimation on data set B

The most accurate estimated densities for data sets A and B are Gaussian and exponential, respectively. While Gaussian and exponential were very effective for their corresponding data set, they were highly inaccurate with the other. The uniform model was consistently ineffective. The Parzen method, while less accurate than the best parametric estimation, was more consistently acceptable.

It is always possible to use the parametric method if the data set is of a known distribution type. However, if the distribution type is not known, a non-parametric method is likely to perform better.

3.0 Model Estimation 2-D Case

A data set of 2D samples containing three classes (red, blue, green) was used for this part of the lab. Firstly, a parametric approach was taken to determine the class boundaries using the Maximum Likelihood method, as seen in figure 3 below.

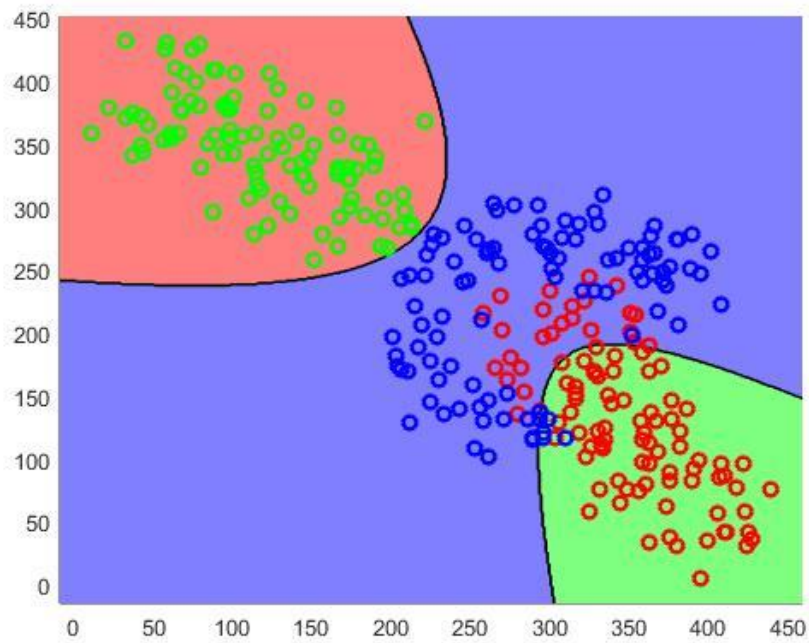


Figure3: Parametric approach to generating class boundaries.

Secondly, a non-parametric approach via a Gaussian Parzen window was taken to determine the class boundaries using the Maximum Likelihood method, as seen in figure 4 below.

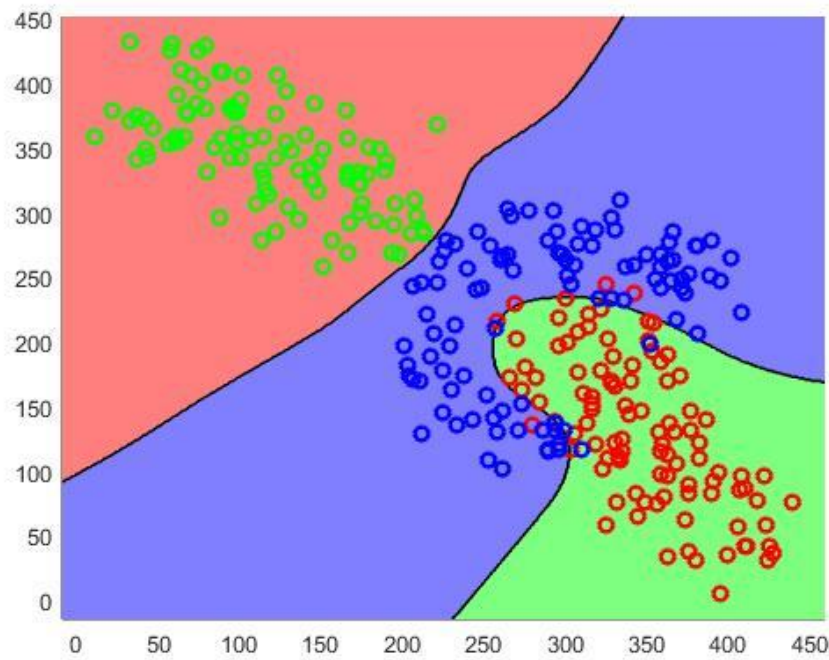


Figure4: Non-parametric approach to generating class boundaries.

4.0 Sequential Discriminants

-

Figure 5: Sequential classifier generated from sample data, iteration 1.

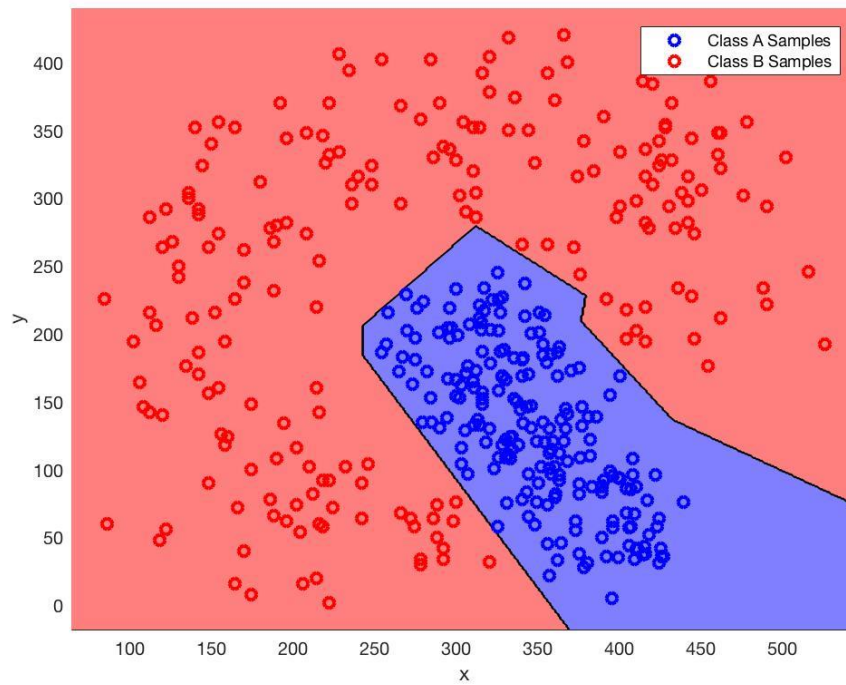


Figure 6: Sequential classifier generated from sample data, iteration 2.

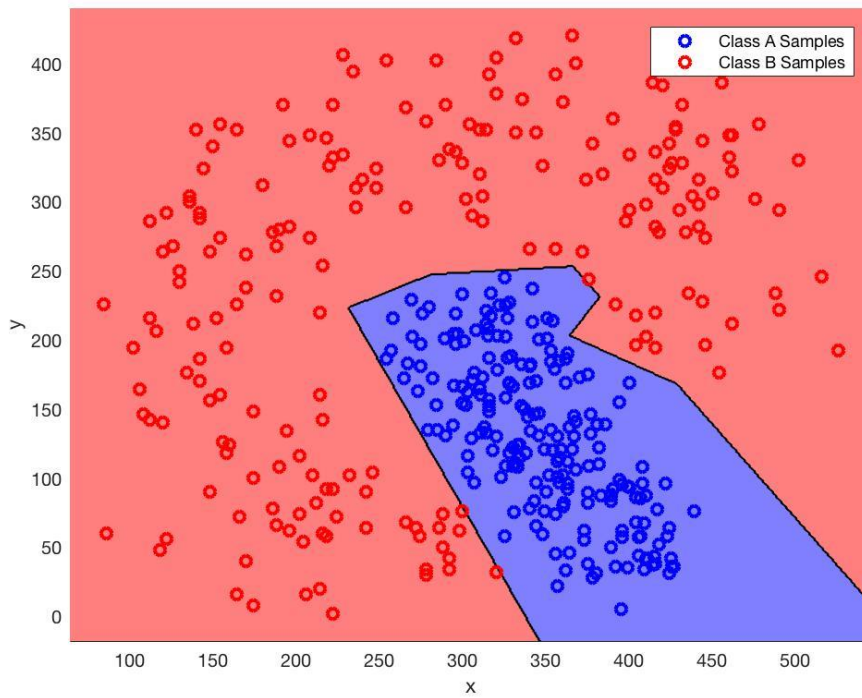


Figure 7: Sequential classifier generated from sample data, iteration 3.

2.

For the classifiers shown in Figures 5 - 7, the training data is completely learnt (overfitted). Discriminants were produced via multiple iterations until all the available training data was exhausted. For every sample in class A, there will exist some discriminant which indicates that the point belongs to class A and that no point in class B was mislabelled as a point in class A. The same is true for every sample of class B. Therefore, if the training data is fed back into the classifier, the error rate will always be **zero**. However, new samples, that were not previously seen, will most likely produce a higher error rate than that of zero.

3.

Figure 8 compares the error rates when the number of discriminants is varied. A single MED discriminant produces the highest error rate. As the amount of discriminants increases, the error rate metrics decrease.

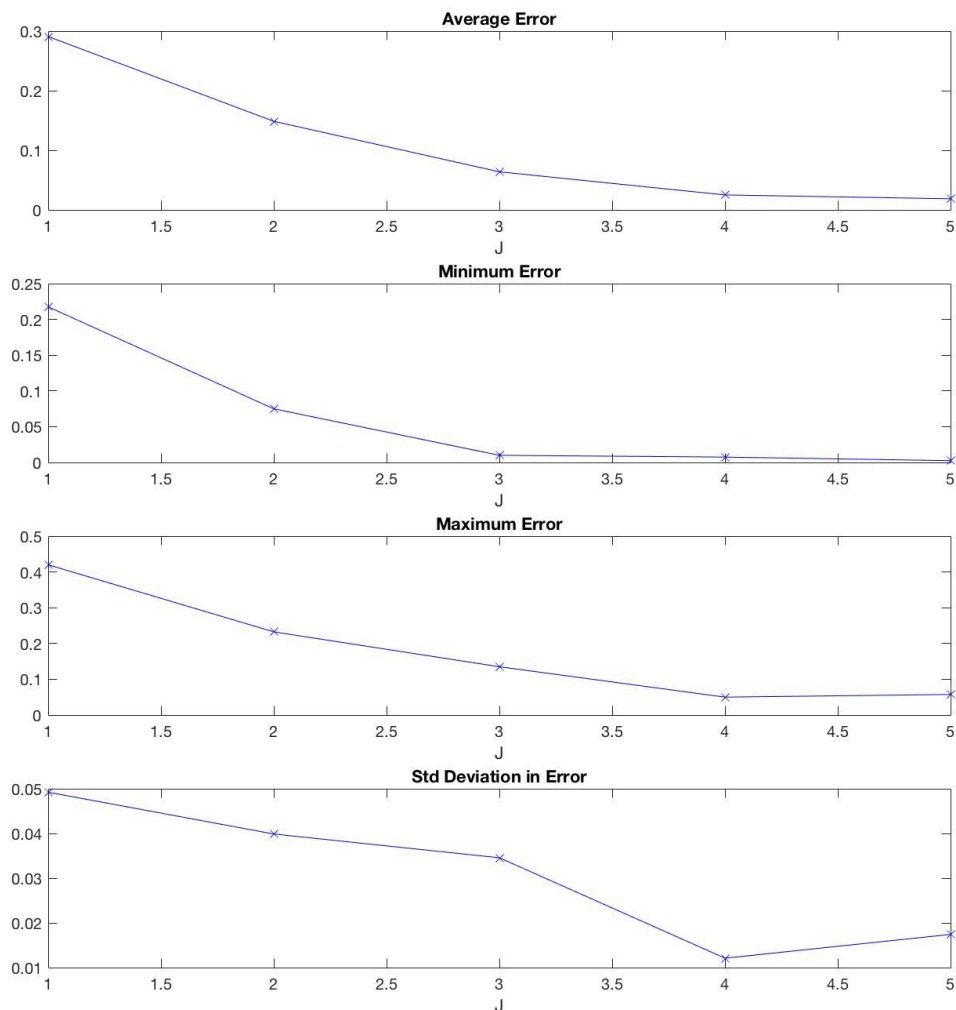


Figure 8: Error rate analysis for varying number of discriminants in classifier.

If the points used for training data were limited in some way, the performance of the classifier would be affected. If all the training point were removed from around one specific region, then chances are that data in that region would be misclassified often. However, if samples were randomly selected from the training data for removal then the classifier is less likely to overfit and would have more generalization for new data, thereby improving its performance.

5.0 Conclusion

The results of section 2 clearly show that parametric estimation is highly effective for data sets with known distributions, while non-parametric estimation is preferred otherwise. This conclusion is reinforced in the 2D case in section 3. Lastly, as can be seen from section 4, combining multiple weak classification techniques can lead to a powerful classifier.