ECE571 Pattern Recognition

Project1: Bayesian Decision Rule

Yang Xu

yxu71@vols.utk.edu

Abstract

This project is applying Bayesian decision rule to classification problem. Three discriminant functions were discussed, and they have very different results. Since decision rule relies on prior probability, I also examined the change of predication with that of prior probability. Finally, bimodal gaussian was implemented and compared to unimodal gaussian. The result showed no dramatic improvement after applying a more complex model.

Introduction

Bayesian was widely in used in real world problem. Different from frequentism, Bayesian rule makes decisions based reasonable assumptions. That's say, prior probability. Introducing prior probability can free us from limitation of samples. This is because, we make estimations by using samples we have. However, we could not know the grand truth. Thus, making decision from frequentism will lead us to the bias contained in the available samples. But decisions based on Bayesian rule can reduce the risk of being biased.

Technical approach

In this project, I used maximum likelihood estimation to estimate gaussian parameters. Discriminant functions were discussed in 3 different cases: 1) case1 estimates parameters by assuming features are statistically independent and they have the same variance; 2) case2 estimates parameters by assuming the covariance matrices for all the classes are identical; 3) case3 estimates parameters by assuming the covariance matrices are different from each category, and decision boundary would be hyper quadratic for 2-D Gaussian. To estimate parameters of bimodal gaussian, density estimation was applied to find means and variances, which I first binned data points and then fit them to density of histogram.

Experiments and results

1. Visualizing training dataset
   The very first step is visualization of training set.

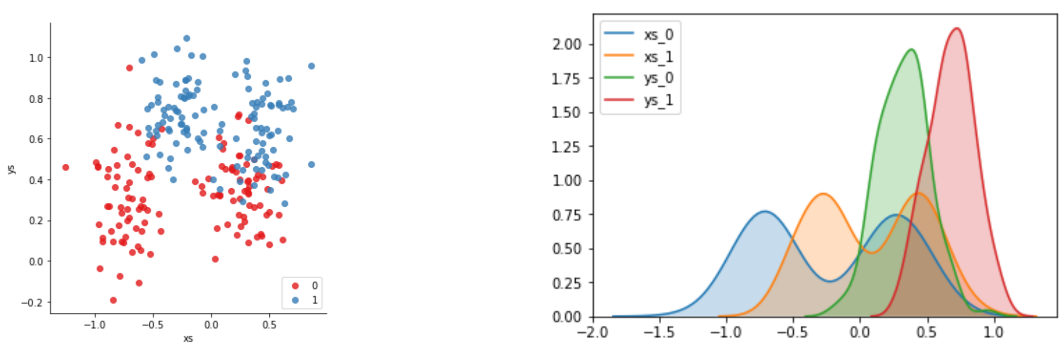


Figure 1

Graphically, two classes are present in this dataset. Class 1 has higher ys mean than class 0. Density of feature xs shows two peaks, indicating bimodal gaussian distribution. However, feature ys only has unique peak and suggests the normal distribution.

2. Decision Boundary
   After implementing the 3 discriminant functions, I got 3 decision boundaries for testing set. Figure 2 shows the 3 decision boundaries. Both case1 and case2 have linear decision boundaries. However, case3 has a quadratic boundary between class 0 and class 1. Different from case1, decision boundary of case2 has smaller slope and has good classification performance (Accuracy showed below).

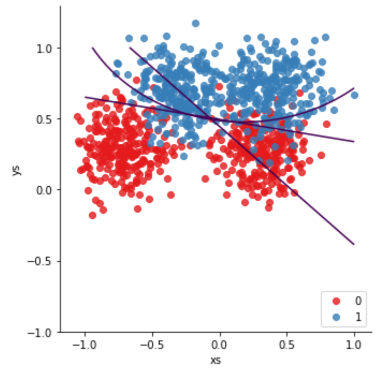


Figure 2

Case1 Accuracy

y_pred	0	1
y_test		
0	340	160
1	127	373

Case2 Accuracy

y_pred	0	1
y_test		
0	451	49
1	54	446

Case3 Accuracy

y_pred	0	1
y_test		
0	471	29
1	86	414

Case1 has poor classification accuracy (0.68 for class0 and 0.746 for class1). Case2 and case3 has similar performance on predication of testing set. The slight difference relies in better accuracy for class1 in case2 and better accuracy for class9 in case3. Even though case2 and case3 has no dramatic performance, case2 is a simpler model with easy implementation.

3. Model Accuracy

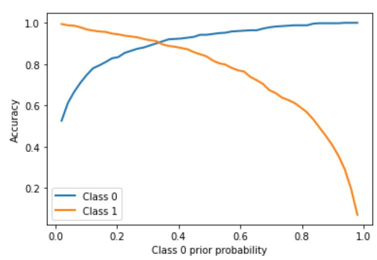
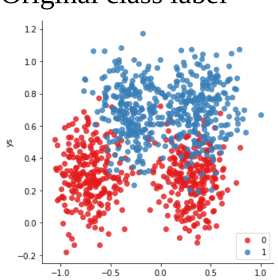


Figure 3

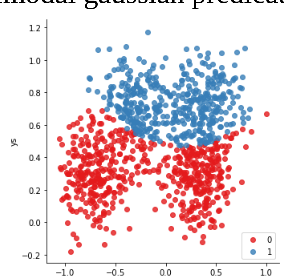
Next, I examined how prior probability affects model accuracy. In this examination, I only discussed case3. As prior probability of class0 increase, predication accuracy for class0 also increases. When the prior probability gets close to 1, the accuracy approaches to 100% accurate. On the opposite, since prior probability of class1 decreases as that of class0 goes up, predication accuracy for class1 decreases, and it dives to 0 when prior probability of class1 drops to 0.

4. Bimodal Gaussian Classification

Original class label



Unimodal gaussian predication



Bimodal gaussian predication

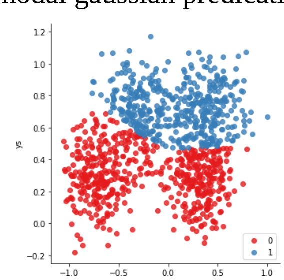


Figure 4

Unimodal Gaussian Accuracy

y_pred	0	1
y_test		
0	469	31
1	78	422

Bimodal Gaussian Accuracy

y_pred_case3	0	1
y_test		
0	471	29
1	86	414

Comparison of unimodal gaussian and bimodal gaussian shows slight difference on classification in this task. Unimodal gaussian has better accuracy on classifying class1 but bimodal gaussian has better accuracy on classifying class0

Discussion

In this project, I implemented Bayesian decision rule for classification on a two-class classification problem, via 3 discriminant functions. Features have correlations to some degree in reality. This is the reason why case1 had poor performance, because it assumes features are statistically independent. Applying complex assumptions greatly improve predication accuracy. Results from case2 and case3 also suggest linear discriminant function can have very good performance if with good assumptions. In fact, most real-life problems are likely to have quadratic complexity, but it can also be solved by linear function, see case2. I examined how prior probability affects predication. Results suggest a balance needed if good predication for each class. Unbalanced prior probability favors predication for heighted class. Supposing that one class has prior probability with 1, there should be no model needed to predict, because of no expectation that the other class will show up. However, this is not true in reality.

Appendix

All implementation is available in the python script and should be run by Python 3.