

CS 534 Machine Learning

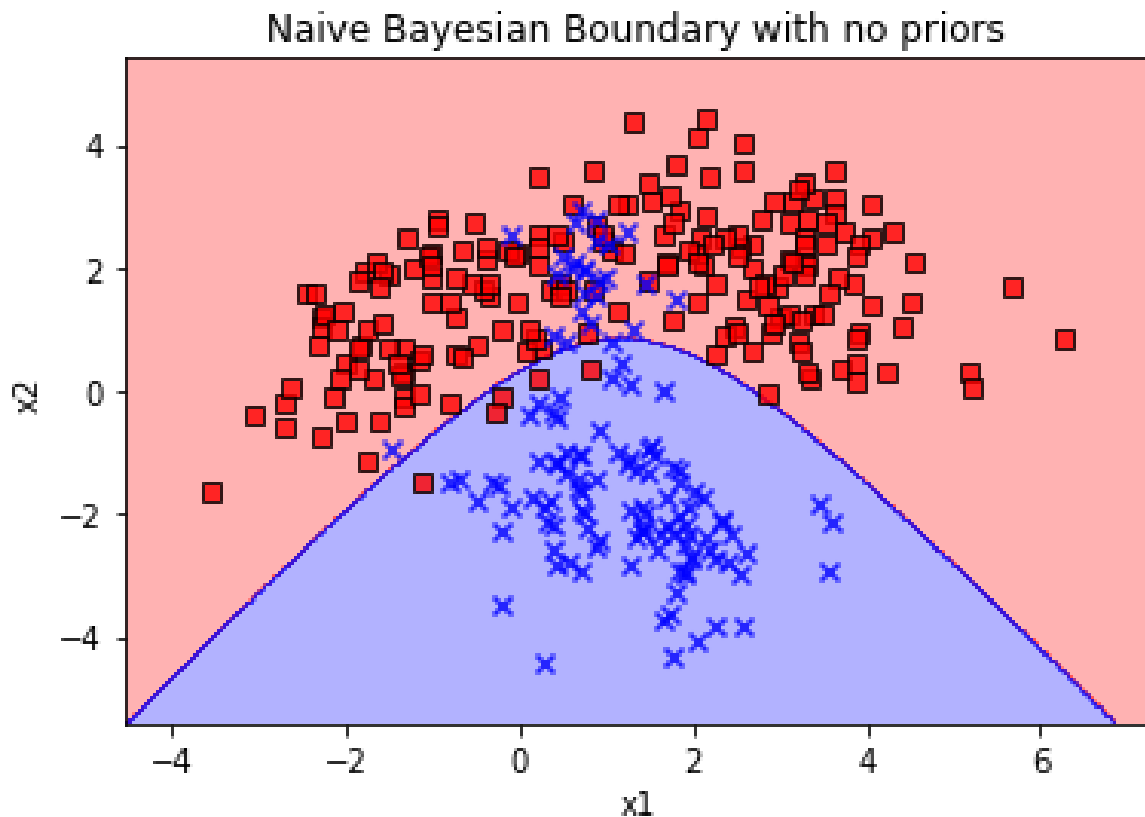
Yicheng (Jason) Wang

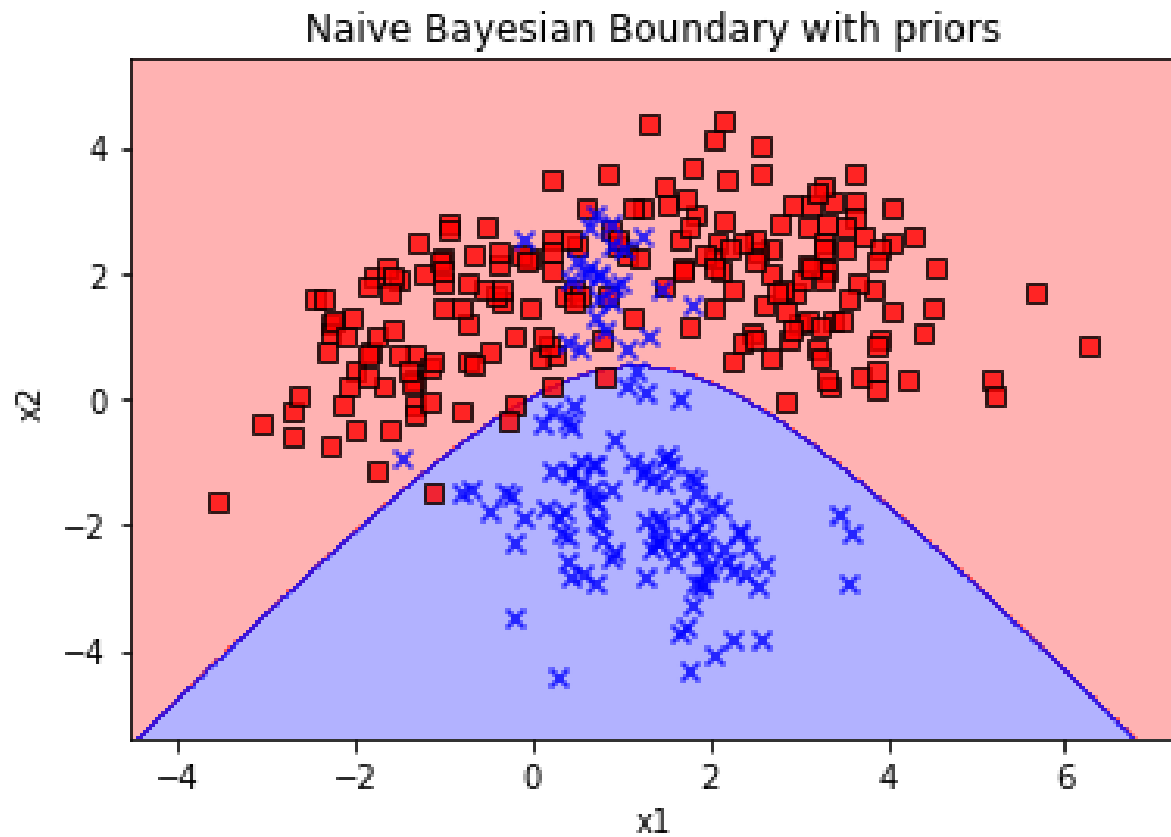
** Please look at the jupyter notebook if you happen to know how to use that, since it contains more explanation and details, the file below is just for briefly and result.

Problem 1:

The mean vector for class 0 is: [1.06242164 1.61910524]
The covariance matrix for class 0 is:
[[4.79170095 0.90180838]
[0.90180838 1.2945715]]

The mean vector for class 1 is: [1.13915258 -1.18380439]
The covariance matrix for class 1 is:
[[0.7560476 -0.5093068]
[-0.5093068 3.19387164]]





Notice there is a tiny difference between the two boundaries.

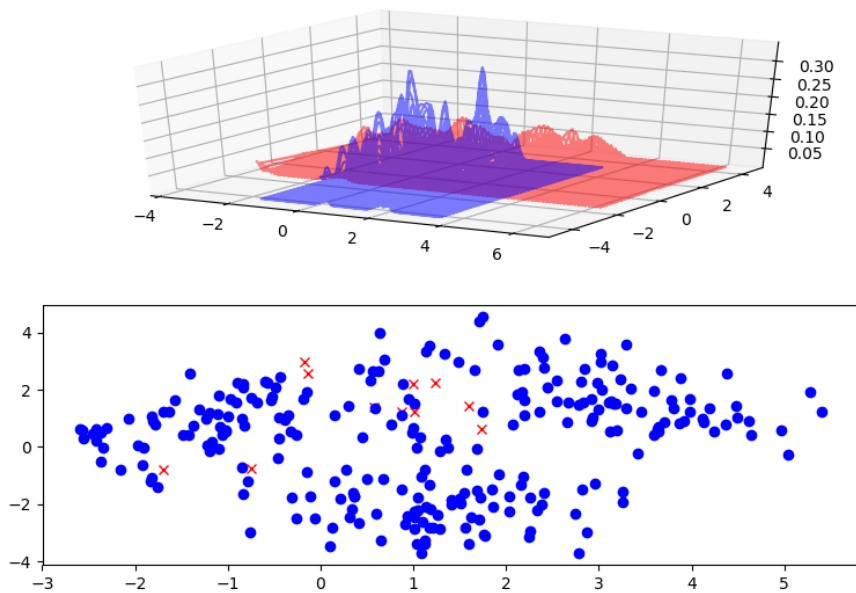
The classification error rate for Bayesian Decision Boundary without priors is: 0.06

The classification error rate for Bayesian Decision Boundary with priors is: 0.06

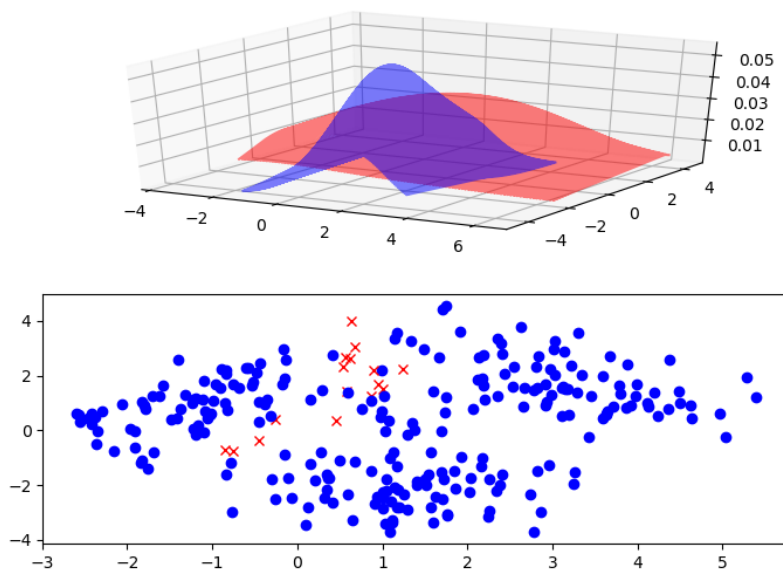
Problem 2:

** In this section, the first graph is the density estimation graph, the second graph is the test data plot, with red cross(x) indicates the misclassified data.

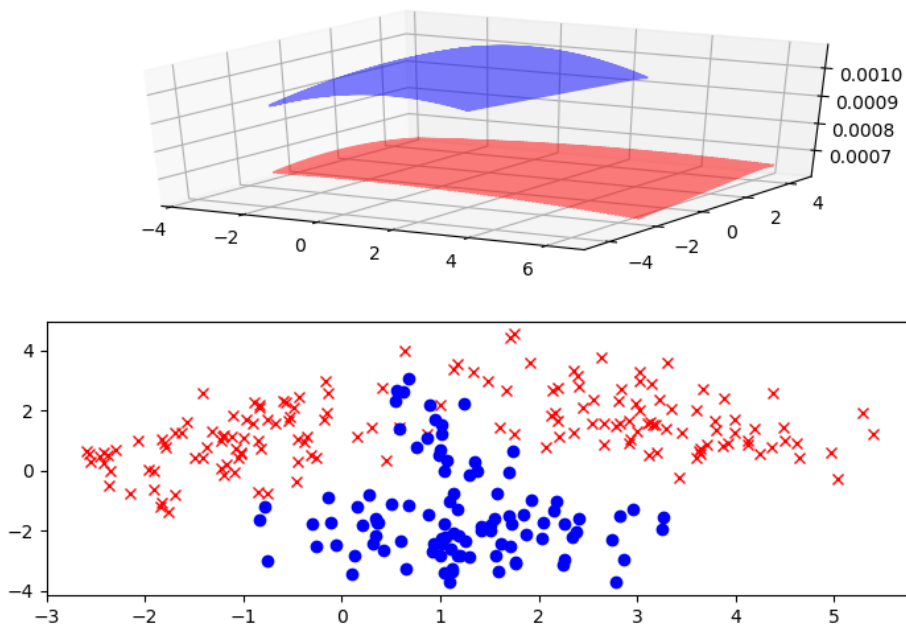
1. Bandwidth = 0.1



2. Bandwidth = 1



3. Bandwidth = 10



The classification error rate for bandwidth = 0.1 is 0.02

The classification error rate for bandwidth = 1 is 0.03

The classification error rate for bandwidth = 10 is 0.31

Problem 3:

When $k = 1$

The sensitivity rate is 0.92

The specificity rate is 0.97

The false discovery rate rate is 0.08

When $k = 5$

The sensitivity rate is 0.96

The specificity rate is 0.98

The false discovery rate rate is 0.04

When $k = 10$

The sensitivity rate is 0.95

The specificity rate is 0.99

The false discovery rate rate is 0.05

K Nearest Neighbour (k = 10)

