

Homework 1 Part B: Practice

Tianyang Chen

1. Half Moon Data Generator and Linear Classifier

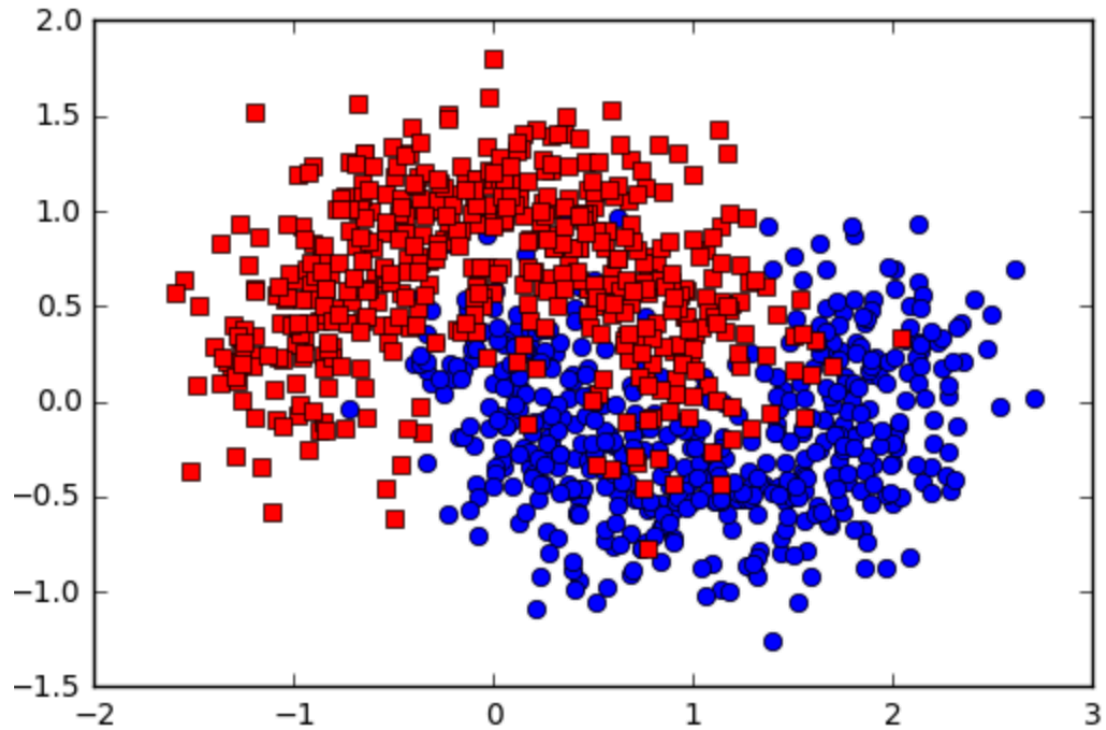


Fig.1 Training Data

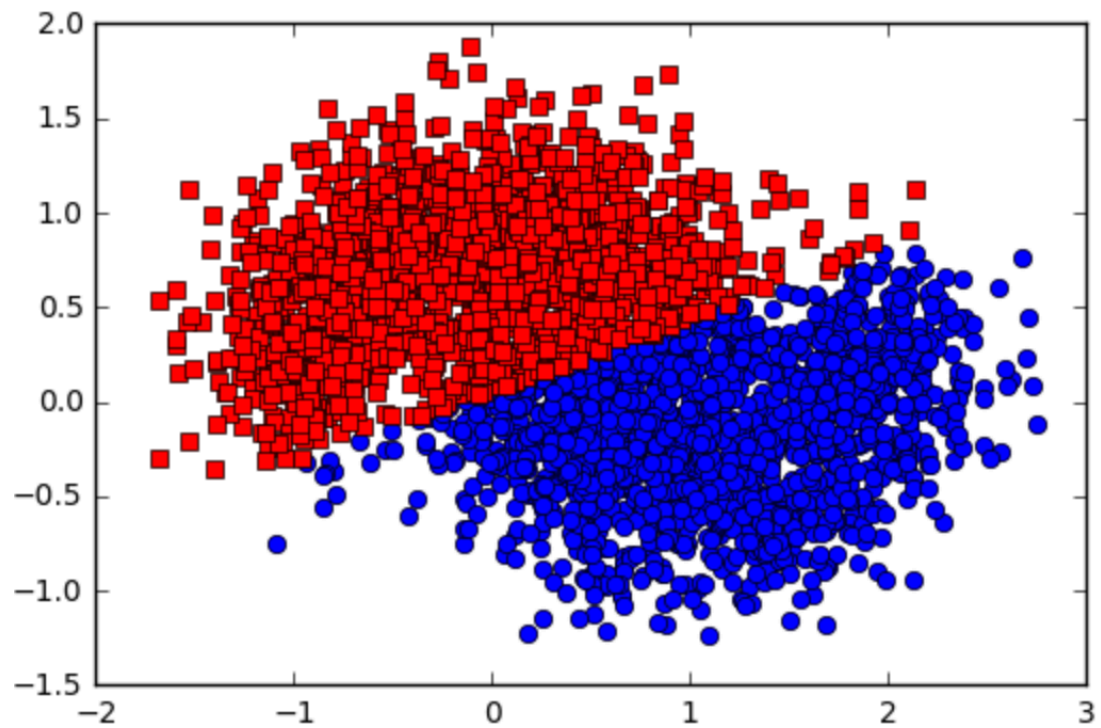


Fig.2 Decision Boundary

2. Naïve Bayes Spam Filter

5-fold cross validation error is 0.197826.

P.S The 5-fold cross validation will be shown in MATLAB's command window after running the program "spam.m". Because the 5 folds were generated randomly, this average 5-fold cross validation may vary a little.

3. Comparing Classifiers

I used the Friedman test to measure the performance of different classifiers. The average rank of these classifiers are shown below.

1	2	3	4	5	6
2.9074	4.7037	4.1852	4.2593	2.4630	2.4815

Fig.3 Average Rank

It seems that classifier 5 and 6 are better than the others.