

HW3 Report

1. Result

```
命令提示符
C:\Users\Xiaobo Zhang\Desktop\project3>python test.py
Data File: funny_data.csv
Train size: 500
ZERO-ONE-LOSS: [0.3651111111111111, 0.3813333333333333, 0.3724444444444447, 0.3742222222222223, 0.3804444444444444, 0.372, 0.3795555555555553, 0.3746666666666665, 0.3748888888888889, 0.3751111111111111]
Mean: 0.374977777778
stdev 0.00452286372736
Baseline error
Mean: 0.403177777778
stdev 0.130218014722
Train size: 2500
ZERO-ONE-LOSS: [0.3556, 0.36, 0.3416, 0.3596, 0.3452, 0.3652, 0.346, 0.3488, 0.3784, 0.3648]
Mean: 0.35652
stdev 0.0107985925009
Baseline error
Mean: 0.4966
stdev 0.114431359338
Train size: 4500
ZERO-ONE-LOSS: [0.366, 0.34, 0.342, 0.356, 0.356, 0.342, 0.32, 0.35, 0.35, 0.362]
Mean: 0.3484
stdev 0.0125475097131
```

```
命令提示符
Mean: 0.4966
stdev 0.114431359338
Train size: 4500
ZERO-ONE-LOSS: [0.366, 0.34, 0.342, 0.356, 0.356, 0.342, 0.32, 0.35, 0.35, 0.362]
Mean: 0.3484
stdev 0.0125475097131
Baseline error
Mean: 0.5618
stdev 0.0892835931177
[[0.3651111111111111, 0.3813333333333333, 0.3724444444444447, 0.3742222222222223, 0.3804444444444444, 0.372, 0.3795555555555553, 0.3746666666666665, 0.3748888888888889, 0.3751111111111111], [0.3556, 0.36, 0.3416, 0.3596, 0.3452, 0.3652, 0.346, 0.3488, 0.3784, 0.3648], [0.366, 0.34, 0.342, 0.356, 0.356, 0.342, 0.32, 0.35, 0.35, 0.362]]
Data File: stars_data.csv
Train size: 500
ZERO-ONE-LOSS: [0.2331111111111111, 0.2693333333333333, 0.2715555555555555, 0.23, 0.2551111111111111, 0.2284444444444445, 0.2486666666666667, 0.2655555555555554, 0.2624444444444444, 0.2362222222222222]
Mean: 0.250044444444
stdev 0.0161333333333
Baseline error
Mean: 0.559377777778
stdev 0.13537703363
```

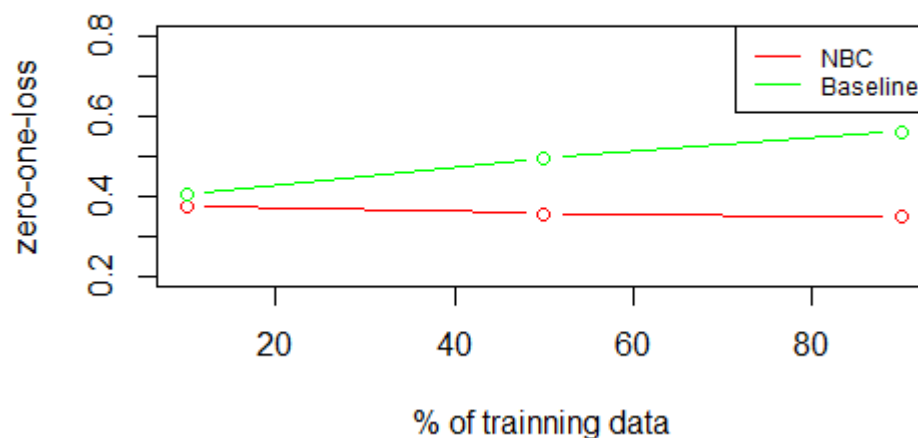
```

命令提示符
Data File: stars_data.csv
Train size: 500
ZERO-ONE-LOSS: [0.2331111111111111, 0.2693333333333333, 0.2715555555555555, 0.23, 0.2551111111111111, 0.2284444444444445, 0.2486666666666667, 0.2655555555555555, 0.2624444444444444, 0.2362222222222222]
Mean: 0.250044444444
stdev 0.0161333333333
Baseline error
Mean: 0.559377777778
stdev 0.13537703363
Train size: 2500
ZERO-ONE-LOSS: [0.22, 0.222, 0.2168, 0.2092, 0.2044, 0.2228, 0.2192, 0.2116, 0.1868, 0.2081]
Mean: 0.21208
stdev 0.0103549794785
Baseline error
Mean: 0.48376
stdev 0.0996787158826
Train size: 4500
ZERO-ONE-LOSS: [0.22, 0.21, 0.2, 0.208, 0.216, 0.212, 0.234, 0.198, 0.202, 0.236]
Mean: 0.2136
stdev 0.0125475097131
Baseline error
Mean: 0.5038

```

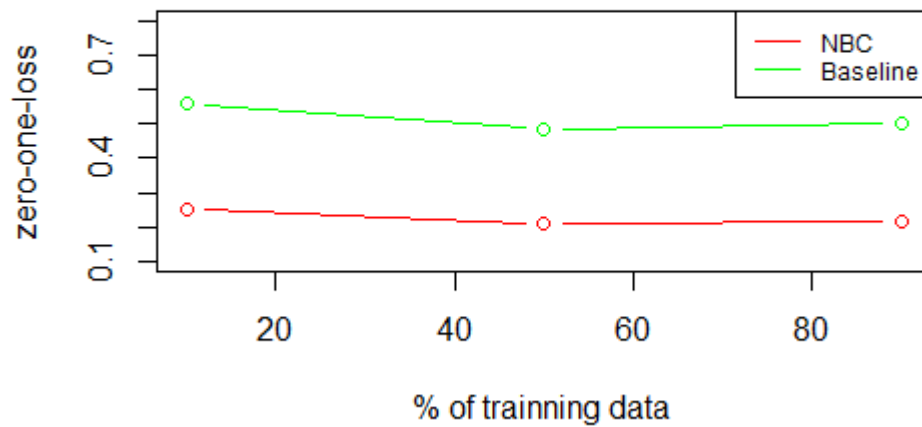
2. Chart
Funny data:

Funny data: zero-one-loss VS training data size



Stars data:

Stars data: zero-one-loss VS training data size



3. Analysis

According both charts, we can conclude that the NBC classifier has much lower zero-one-loss error than base line error. We can find out that when the data size larger, the error rate of NBC will become much slower. The zero-one-loss goes lower when data size from 10% to 50%, but the error doesn't change a lot when the data size from 50% to 90% for NBC classifier.