**Assignment 1**

Min Zhang , [mz86@njit.edu](mailto:mz86@njit.edu), 31172077

1. Evaluating a classification algorithm

b)

|  |  |  |  |
| --- | --- | --- | --- |
| Output  Labels | Ture Labels | | |
|  | 1 | -1 |
| 1 | 2(TP) | 5(FP) |
| -1 | 4(FN) | 3(TN) |

E=(FP+FN)/(TP+FP+FN+TN)=0.6428

BER=0.5\*((FP/FP+TN)+(FN/FN+TP))=0.6458

Precision=TP/TP+FP=0.2857

Recall=TP/TP+FN=0.3333

c)

example : to classify 2 group one with cancer and one without.

If a classifier can perform with 90% precision, but 90 percent of the population does not have cancer and the 10% that do are misclassified by the classifier. In this case, each one of measures does not reflect to power of classification.

Balanced Error generally works better

1. Nearest means vs Naïve Bayes.

e)

Nearest means results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | E | BER | Precision | Recall |
| Breast\_cancer | 0.10896 | 0.14143 | 0.86063 | 0.98599 |
| Qsar | 0.33838 | 0.33117 | 0.80427 | 0.64663 |
| Ionosphere | 0.26495 | 0.26253 | 0.60645 | 0.74603 |
| Climate | 0.12037 | 0.0855 | 0.99538 | 0.87246 |

Naïve Bayes Results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | E | BER | Precision | Recall |
| Breast\_cancer | 0.07557 | 0.06405 | 0.98757 | 0.89075 |
| Qsar | 0.18104 | 0.22552 | 0.83159 | 0.9113 |
| Ionosphere | 0.2735 | 0.24825 | 0.58241 | 0.84126 |
| Climate | = 0.03703 | 0.1681 | 0.97023 | 0.98987 |

Naïve Bayes has the higher accuracy

f) I increased all the variance by 1. Variance is zero means this feature (column) remain the same across all vector. The algorithm is trying to find the vectors that are least likely to be random distribution based on the properties of a certain class. I think changing all the variance by an equal amount would have minimum effect on the comparison.