

**Jiyang Liu 4731134**

文本

描述已自动生成



**1a.**

Use the binomial distribution:

For m=1, p=0.04,

For m=1, p=0.75,

For m=1000, p=0.04,

For m=1000, p=0.75,

For m=1000000, p=0.04,

For m=1000000, p=0.04,

|  |  |  |  |
| --- | --- | --- | --- |
|  | m=1 | m=1,000 | m=1,000,000 |
| p=0.04 |  |  |  |
| p=0.75 |  |  |  |

文本, 信件

描述已自动生成

**1b.**

The two coins are the same, therefore,

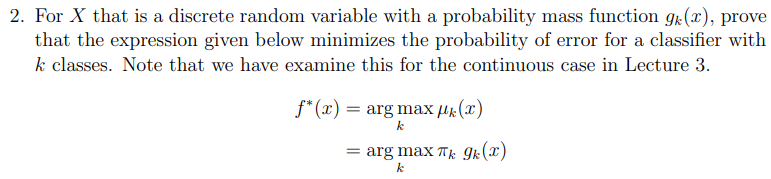
To calculate the probability,

Where , ,

**The specific calculation code is in the HW1.ipynb.**

图表, 折线图

描述已自动生成



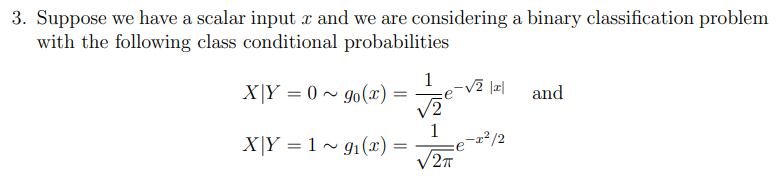
Consider an arbitrary classifier f and denote the decision regions

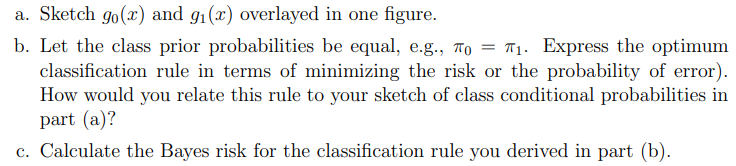
To minimize the probability of error for a classifier with k classes,

To maximize this, we should select such that

Therefore, the optimal

Or equivalently





**3a. The specific calculation code is in the HW1.ipynb**

**图表, 折线图

描述已自动生成**

**3b.**

The prior probabilities are equal, which is,

The optimal classification rule,

If is greater than , then the class would be . Otherwise, the class would be .

In the sketch when is on the top then the class would be . Otherwise, the class would be .

**3c.**

We get,

Therefore, the decision region is,

Then,

文本

描述已自动生成

In the linear classifier we have,

Where the Mahalanobis distance is,

Substitute and move terms, we get,

Notice that the dimensions of is,

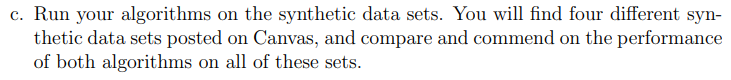
Which is a single value, therefore,

Substitute and normalized, we get,

文本

描述已自动生成

**5a, 5b The specific algorithm code is in the HW1.ipynb**



**The validation and data visualization code are in HW1.ipynb**

**Synthetic1:**

**图表, 散点图

描述已自动生成**

文本

描述已自动生成

**Discussion:**

In Synthetic1, there is basically no outlier in the data, and the similar points of the two classifications are relatively concentrated, and the distance between the central areas is very large. Both classifiers completed the classification perfectly, with a risk of 0.

**Synthetic2:**

**图表, 散点图

描述已自动生成**

文本

描述已自动生成

**Discussion:**

In Synthetic2, it is similar to Synthetic1, but with a small number of outliers. Both classifiers have a small risk, and the risk of the simple classifier is slightly smaller.

**Synthetic3:**

**图表, 散点图

描述已自动生成**

文本

描述已自动生成

**Discussion:**

In Synthetic3, the two types of data are mixed together and cannot be separated well, so the risk of the classifier is much higher. The risk of simple classifiers is higher than that of ordinary classifiers.

**Synthetic4:**

**图表, 散点图

描述已自动生成**

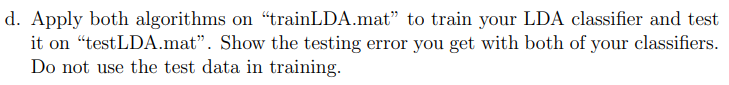
文本

描述已自动生成

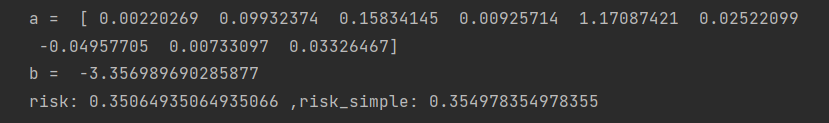
**Discussion:**

In Synthetic4, the two types of data are mixed to the highest degree, so its risk is the highest. The risk of simple classifiers is higher than that of ordinary classifiers.

**In general, simple classifiers perform better when there are fewer outliers in the data; when the degree of data mixing is high, the risk of simple classifiers is higher, but not much worse, and the calculation of simple classifiers The amount should be less. The reason may be that the simple classifier only takes the trace of the deviation matrix, so its classification is more general and its ability to resist noise is stronger.**



**Training:**



**Testing:**

图形用户界面, 文本

中度可信度描述已自动生成