

**** 1

**模式识别与机器学习**

**Synthetical Design of Bayesian Classifier**

实验报告

|  |  |
| --- | --- |
| 院（系）名称 | 自动化科学与电气工程学院 |
| 专业名称 | 自动化 |
| 学生姓名 | 伍翌嘉 |
| 学号 | 15031131 |
| 任课老师 | 李阳、秦曾昌 |

2018年5月4日

**Experiment 2: Synthetical Design of Bayesian Classifier**

**2.1 Introduction**

Linear perceptrons allow us to learn a decision boundary that would separate two classes. They are very effective when there are only two classes, and they are well separated. Such classifiers are referred to as discriminative classifiers. In contrast, generative classifiers consider each sample as a random feature vector, and explicitly model each class by their distribution or density functions. To carry out the classification, the likelihood function should be computed for a given sample which belongs to one of candidate classes so as to assign the sample to the class that is most likely. In other words, we need to compute  for each class  However, the density functions provide only the likelihood of seeing a particular sample, given that the sample belongs to a specific class. i.e., the density functions can be provided as . The Bayesian rule provides us with an approach to compute the likelihood of the class for a given sample, from the density functions and related information.

**2.2 Principle and Theory**

The essence of the Bayesian approach is to provide a mathematical rule explaining how you should change your existing beliefs in the light of new evidence. In other words, it allows us to combine new data with their existing knowledge or expertise. The canonical example is to imagine that a precocious newborn observes his first sunset, and wonders whether the sun will rise again or not. He assigns equal prior probabilities to both possible outcomes, and represents this by placing one white and one black marble into a bag. The following day, when the sun rises, the child places another white marble in the bag. The probability that a marble plucked randomly from the bag will be white (i.e., the child’s degree of belief in future sunrises) has thus gone from a half to two-thirds. After sunrise the next day, the child adds another white marble, and the probability (and thus the degree of belief) goes from two-thirds to three-quarters. And so on. Gradually, the initial belief that the sun is just as likely as not to rise each morning is modified to become a near-certainty that the sun will always rise. In terms of classification, the Bayesian theorem allows us to combine prior probabilities, along with observed evidence to arrive at the posterior probability. More or less, conditional probabilities represent the probability of an event occurring given evidence. According to the Bayesian Theorem, if and X are known or given, the posterior probability can be derived as follows



Let the series of decision actions as , the conditional risk of decision action  can be computed by



Thus the minimum risk Bayesian decision can be found as



**2.3 Objective**

The goals of the experiment are as follows:

(1) To understand the computation of likelihood of a class, given a sample.

(2) To understand the use of density/distribution functions to model a class.

(3) To understand the effect of prior probabilities in Bayesian classification.

(4) To understand how two (or more) density functions interact in the feature space to decide a decision boundary between classes.

(5) To understand how the decision boundary varies based on the nature of density functions.

**2.4 Contents and Procedure**

**Stage 1: Bayesian classifier for the classification of two classes of patterns**

The dataset of observed values and prior probabilities are provided as follows:

Ω1={ -3.9847，-3.5549，-1.2401，-0.9780，-0.7932，-2.8531，-2.7605，-3.7287，

-3.5414，-2.2692，-3.4549，-3.0752，-3.9934， -0.9780，-1.5799，-1.4885，

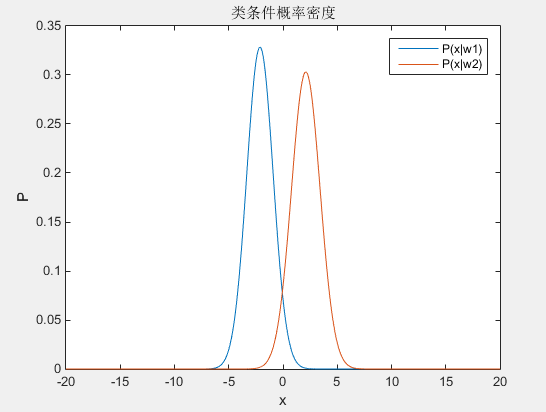
-0.7431，-0.4221，-1.1186，-2.3462，-1.0826，-3.4196，-1.3193，-0.8367，

-0.6579，-2.9683}

Ω2= { 2.8792， 0.7932，1.1882，3.0682，4.2532，0.3271,0.9846,2.7648,2.6588}



As we know that the conditional probability distributions are Gaussian, the first thing is to fit conditional probability density function of these two samples as shown in Fig.1. I draw the curves by calculating the mean and variance of two datasets and put them into Gaussian formula.

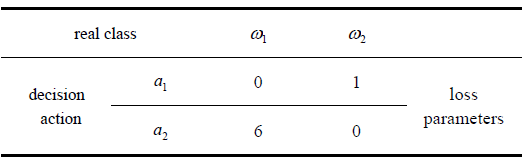


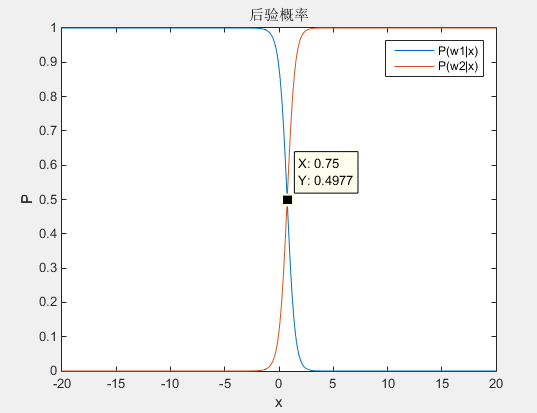
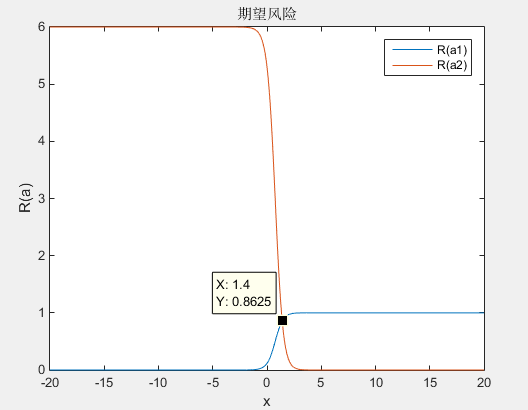
**Fig.1 conditional probability density**

Then posterior probability density functions can be calculated with Bayesian function as shown in Fig.2. From these function, we can make minimum error decision when we don’t know the loss parameters. The decision boundary is about equal to 0.75.

Considering decision loss parameters, minimum risk decision can be made by calculating the risk of two samples as shown in Fig.3. The decision boundary is about equal to 1.4.

**Table 1 the loss parameters for different decision**



**Fig.2 posterior probability density Fig.3 risk distribution**

Comparing the minimum error decision and minimum risk decision, we can easily find that the decision boundary is different. In the minimum risk decision, we are more likely to choose  because of the synthesis consideration of loss.

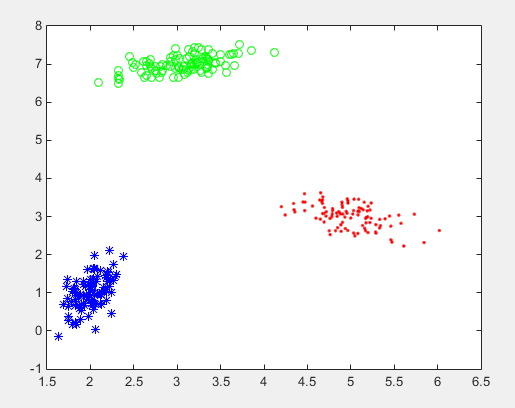
**Stage 2: Bayesian classifier for 2-dimention 3-class dataset**

1. The first step is to create a data set. In order to observe more situation, a random dataset would be better, so I designed function priorP(), lossMatrix() and createdata() to create required data for Bayesian classifier. The mean feature value of 3 classes are decided by input value in order to make sure that they have obvious different features. To get a better display of results, I choose to create 2-dimension 3-class datasets. Fig.4 shows the distribution of 3 classes random data.

The related randomly generated prior probability are P(w1)=0.6, P(w2)=0.3, P(w3)=0.1, and the loss matrix are as Table 2.

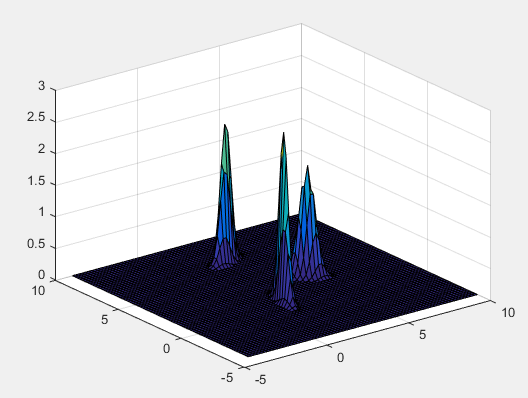
**Table 2 the random loss matrix**

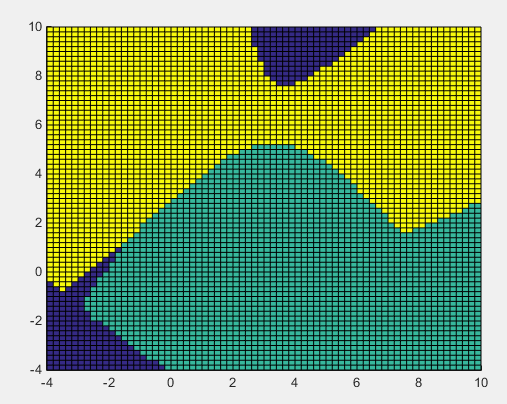
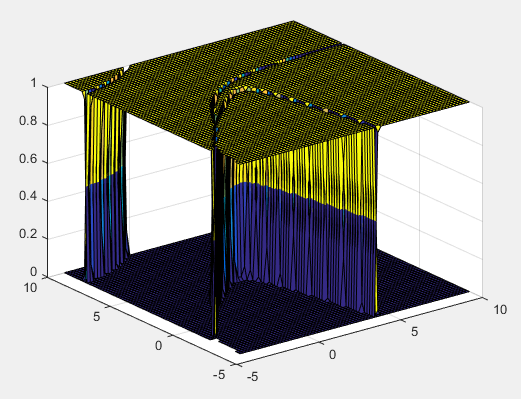
|  |  |  |  |
| --- | --- | --- | --- |
| Real class | W1 | W2 | W3 |
| α1 | 0 | 3 | 9 |
| α2 | 6 | 0 | 2 |
| α3 | 8 | 4 | 0 |



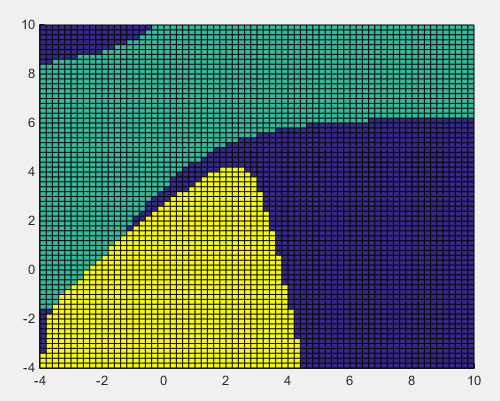
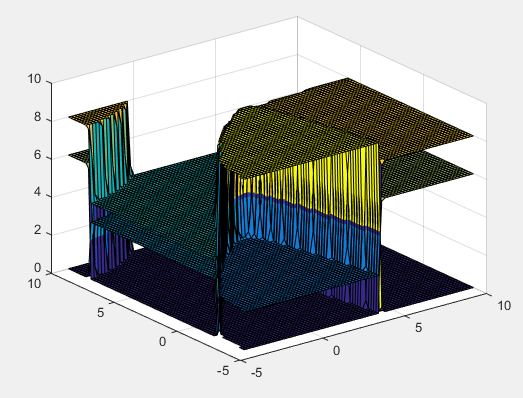
**Fig.4 2-dimension 3-class data**

1. The conditional probability density, posterior probability density and risk distribution condition are calculated by the Bayesian classifier and shown as Fig.5, Fig.6 and Fig.7, separately. We can see that the final decision boundary of two kinds of decision strategy are quite different from each other, more obvious than the first 2 class dataset example. In this experiment, the importance of the choose of decision strategy and parameters of loss are reflected, which may make the final decision vary a lot.



**Fig.5 conditional probability density** 

**Fig.6 posterior probability density**



**Fig.7 risk distribution**

**Stage 3: intrinsic relationship between the classifier of two classes and the one of multiple classes**

The principal, calculating process and algorithm structure of the classifier of two and multiple classes are almost the same. The only difference is that we need to compare data of more classes after getting the posterior probability density and risk distribution.

For the increase of number of features, difference are the calculation complexity is higher in multiple classes because it changes from one-dimentional operation to matrix operation, which may be even complex in higher dimension.

**5 Experiences**

After conducting this experiment, I learn to use Bayesian classifier to make minimum error and minimum risk decision. The difference of two decision boundary in the second experiment impressed me a lot, which make me know the influence of loss matrix and some other parameters. In addition, Bayesian is a nice example of generative classifiers which can separate non-linear separable data by different easily understand strategies. Compare to perceptron learning, such classifiers are more useful in analyzing common data.

**Appendix**

The link of electronic version and codes of this project is:

*https://github.com/sixer51/Pattern-Recognition-Machine-Learning*