成绩

北京航空航天大學
BEIHANG UNIVERSITY

Experiments for “Pattern Recognition and Machine Learning”

Experiment 2
Synthetical Design of Bayesian Classifier

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 学 号 15071135

学 生 姓 名 刘雨鑫

2018 年 5 月 4 日

2.1 Introduction

Linear perceptron allows 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 $p(\omega_i|X)$ for each class ω_i. 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 $p(X|\omega_i)$. 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 $P(\omega_i)$, $p(X|\omega_i), i = 1,2,\ldots,c.$ and X are known or given, the posterior probability can be derived as follows,

$$p(\omega_i|X) = \frac{p(X|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(X|\omega_j)P(\omega_j)} \quad i = 1,2,\ldots,c\#(1)$$

Let the series of decision actions as $\{a_1,a_2,\ldots,a_c\}$, the conditional risk of decision action a_i can be computed by

$$R(a_i|X) = \sum_{j=1,j\neq i}^{c} \lambda(a_i,\omega_j)\, P(\omega_j|X) \quad i = 1,2,\ldots,c\#(2)$$

Thus, the minimum risk Bayesian decision can be found as

$$a_k^* = Argmin_i\ R(a_i|X) \quad i = 1,2,\ldots,c\#(3)$$

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:

First of all, I calculate the mean and variance of the given samples by the program I wrote, then I set the mean and variance as the μ and σ of the Gaussian normal distribution. The conditional probability is shown in the fig.1.

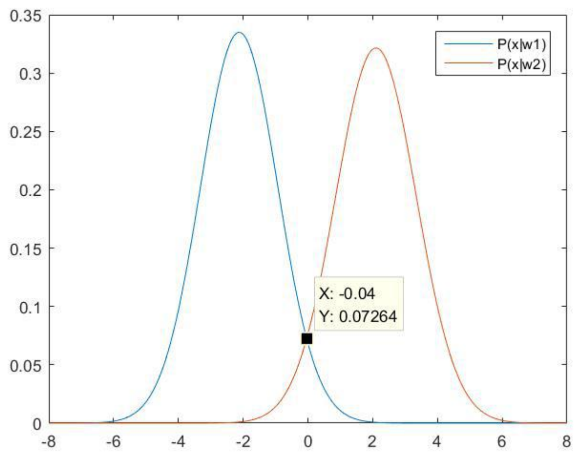

Fig.1. conditional probability of w_1 and w_2

Next, I worked out the posterior probability by the Bayesian Theorem, and it is shown in the fig.2.

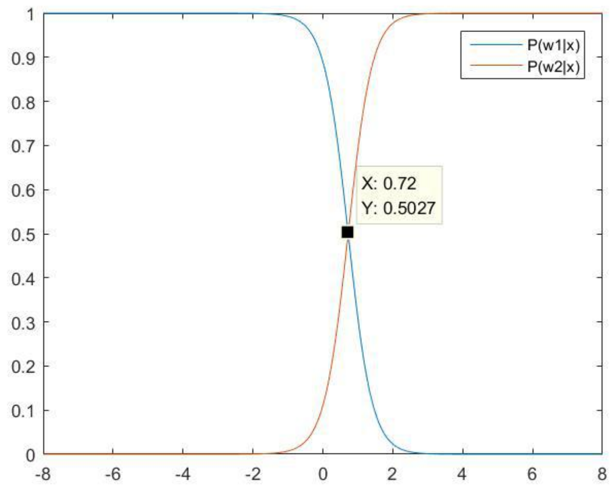

Fig.2. posterior probability of w_1 and w_2

By the program I found the classifying decision boundary without considering decision loss is 0.72, which means the sample that is small than -0.04 will belong to w_1, else belong to w_2.

Then I program the minimum probability based on the decision table 1, and the result is shown in the fig.3.

Table 1 the loss parameters for different decision

real class		ω_1	ω_2	
decision action	a_1	0	1	loss parameters
	a_2	6	0	

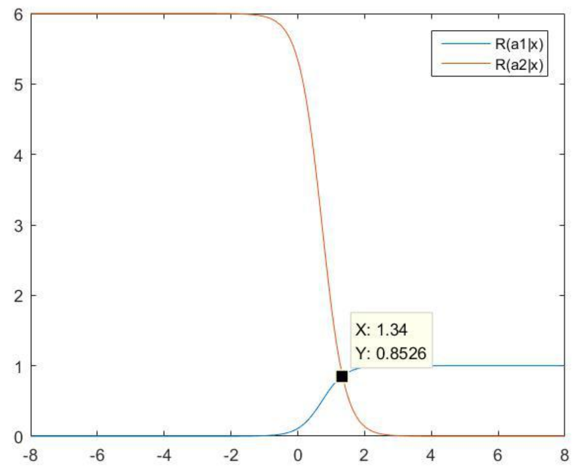

Fig.3. risk probability of w_1 *and* w_2

By the program I found the classifying decision boundary with minimum risk is 1.34, which means the sample that is small than 1.34 will belong to w_1, else belong to w_2.

Stage 2:

I create three classes of samples from normal distributions that can be classified, it is like in the fig.4. Then I set the prior probability P(w_1) = 0.6, P(w_2) = 0.3,P(w_1) = 0.1.

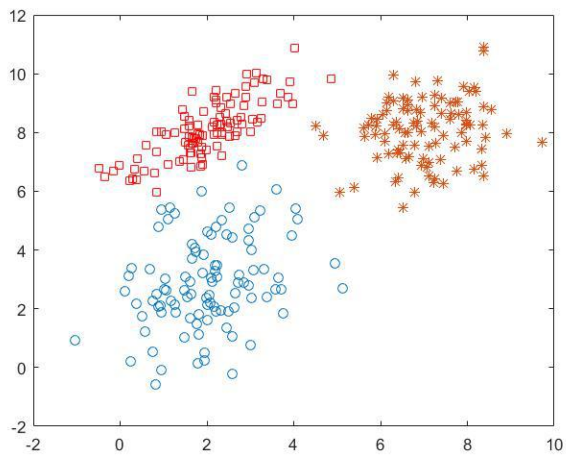

Fig.4. three classes of samples

Next, I work out the conditional probability which is like in stage 1 is shown in the fig.5.

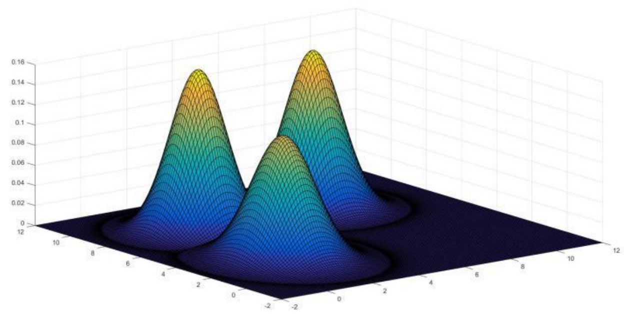

Fig.5. conditional probability of w_1, w_2 *and* w_3

Then I work out the posterior probability by the Bayesian Theorem which is like in stage 1, and it is shown in the fig.6.

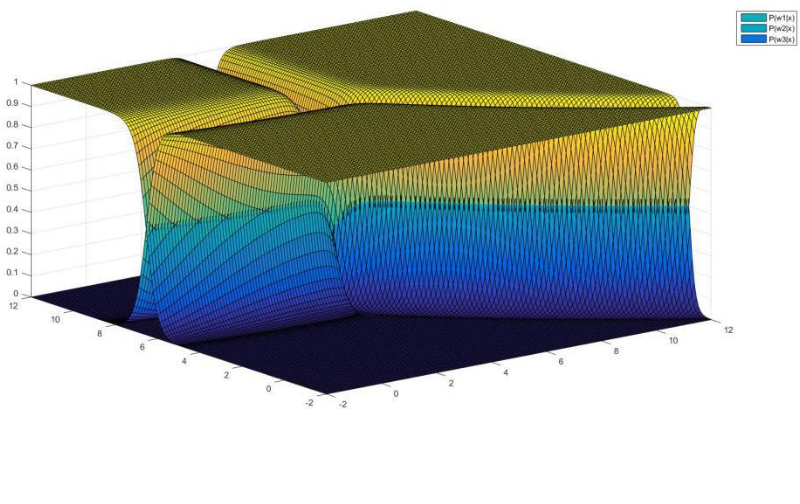

Fig.6. posterior probability of *w_1, w_2 and w_3*

Next I determine the classifying decision boundary without considering decision loss by the program and the result is shown in the fig.7.

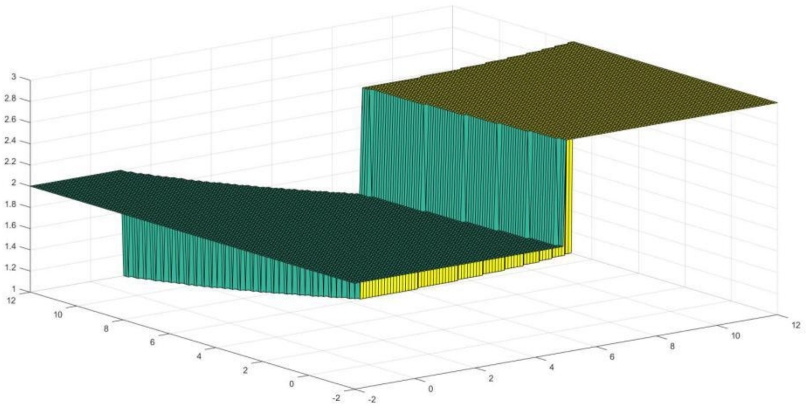

Fig.7. the classifying decision boundary without considering decision loss

Then I program the risk probability based on the decision table 2, and the result is shown in the fig.8.

Table 2. the loss parameters for different decision

Real class		ω_1	ω_2	ω_3	
Decision action	a_1	0	3	1	Loss parameters
	a_2	2	0	1	
	a_3	3	6	0	

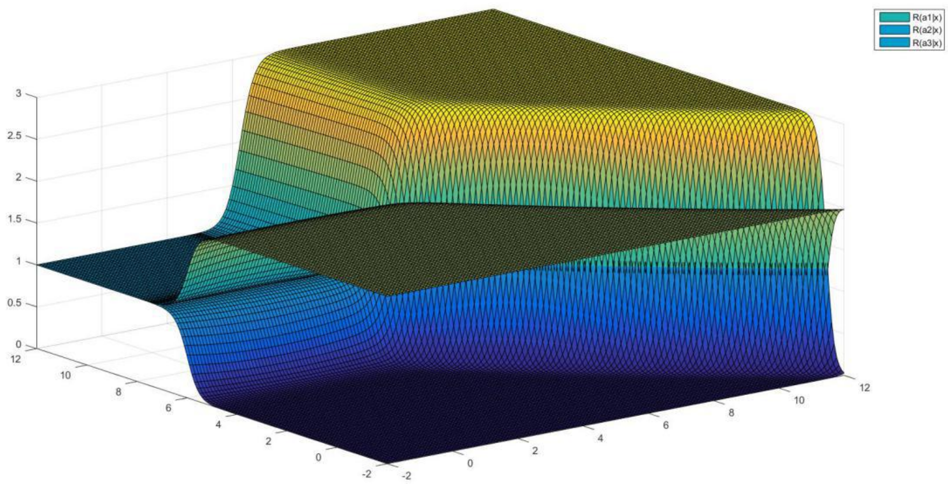

Fig.8. risk probability of *w_1, w_2 and w_3*

Next I determine the classifying decision boundary with minimum risk by the program and the result is shown in the fig.9.

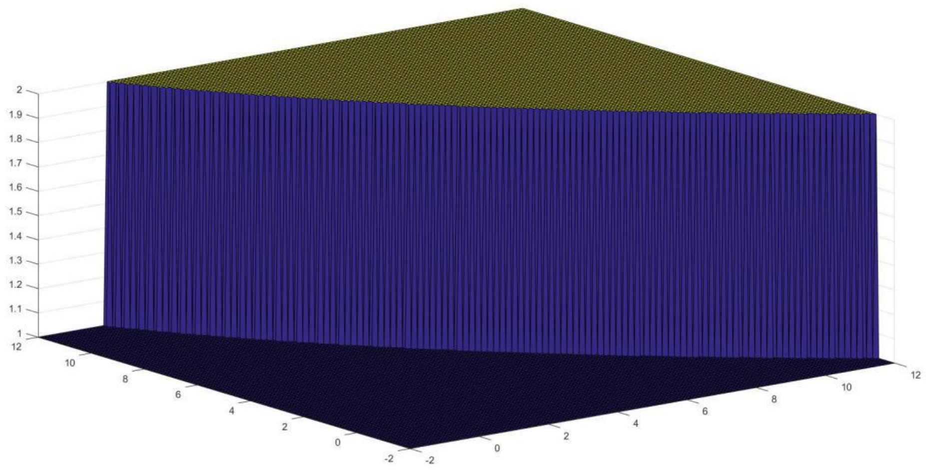

Fig.9. the classifying decision boundary with minimum risk

Through the comparison of the two kinds of experiments above, I think that the basic theory and the formula is the same, but for multiple classes many variable probabilities require more complex calculations, and regional spatial requirement setting and calculation.

2.5 Experience

Through this experiment, I have fully understood the principle of Bayesian classifier, and how to program to achieve this function, learn that Bayes classifier principle is not difficult and easy to achieve. Additionally I also have a deeper understanding of the concept of dimension.

2.6 Code

The code can be download at
https://github.com/Jackboomer/Experiments-for-Pattern-Recognition-and-Machine-Learning.git .