Jared Yu

Data Mining

Module 3 Assignment

1. Let the conditional densities for a two-category one-dimensional problem be given by the Cauchy distribution
   1. If , show that if . Sketch for the case , , . How does behave as ? as ?

Ans:

The first goal is to show that if .

It is given also that .

So, it has been shown that when , then . This works when .

The next goal is to sketch for the case , , . The formula for is as follows:

It is possible to simplify it further, but it was calculated in RStudio. Therefore, each of the components (i.e., likelihood, prior, and posterior) were calculated separately and combined into a single function. Then, the result was calculated for a series of -values between .

Below in Figure 1 is a plot for the case , , . It is assumed that . The range of -values is between . It can be seen that as , starts to approach 0.5.

A screenshot of a cell phone

Description automatically generated

Figure 1 The above figure shows a plot of against for . It can be seen that the probability converges towards 0.5 as .

* 1. Using the conditional densities in part a, and assuming equal *a priori* probabilities, show that the minimum probability of error is given by

Sketch this as a function of .

Ans: Reference ([1])

From part a we have that is the decision boundary, since at that point . Furthermore, we can state that we’d classify a point as if . Using this inequality, it would also imply that for , we’d classify the point as belonging to if .

The reasoning is as follows:

Without loss of generality, let us assume that (we know does not work from part a). This helps to prevent exploring too many cases associated with the inequality. From there we have:

The average probability of error is as follows:

The reason is that probably of the error given the data is as follows,

Then, do a change of variables and .

Also, since arctangent is an odd function, . Here, we have that and so is the negative of . Therefore, .

Going back to , if , then the rule for classifying would be lead to . The result of this is that . Thus, it follows then that,

The sketch of this as a function of can be seen below.

A close up of a map

Description automatically generated

Figure 2 The above plot shows as a function of .

1. The Poisson distribution for discrete , and real parameter is
   1. Find the mean of .

Ans: Reference ([2])

Next, consider the change of variables, .

In the end, the expected value was shown to be multiplied by the sum over all possible values of an alternate Poisson probability mass function () which equates to 1.

* 1. Find the variance of .

Ans: Reference ([3])

First, let us solve for :

Next, consider the change of variables, .

In the end, the expected value was shown to be multiplied by the sum over all possible values of an alternate Poisson probability mass function () which equates to 1. Therefore, , since it has been shown already that .

* 1. Find the mode of .

Ans: (Reference: [4])

There are two possibilities, whether is an integer or a fraction. In the case of being an integer, the probability mass function (pmf) has a bimodal distribution and therefore two distinct modes. In the case of being a fraction, then the pmf has a unimodal distribution and therefore only one mode.

First, let us consider the case when is an integer. Let us look then at the pmf itself,

The mode is the value of that maximizes the probability mass of this pmf. Looking at the pmf, it is apparent that the “” term remains constant and does not vary depending on . It will first be noted though that the values of that maximize are and . When is equal to either of these values, they are equivalent. This will be shown as follows:

To show that this is the maximum probability for , we will consider two alternate values of , and .

Case 1:

In the above situation, the denominator is obviously larger than in the situation before with , therefore it is a smaller number overall. This would be the case for any integer or larger.

Case 2:

Again, like with , the denominator is larger than before with , therefore it is also a smaller overall value. This would apply for any integer smaller than . Therefore, it has been shown that in the case of being an integer, the mode is both and .

Next, let us consider the alternate situation when is a fraction. The Poisson distribution is a discrete density function and so it can only take on values that are integers. Therefore, the mode of it must also be an integer. So, although itself is possibly a fraction, the mode must be an integer.

We will begin by looking at the pmf again carefully,

Once again, the “” term does not depend on , and so it is not directly important for finding the value that maximizes the probability for a given (arbitrary) . So, again we focus on the term . In the numerator, it shows “” which is always increasing as increases. However, the denominator “” is also increasing as increases. The goal then is to find the integer at which this is maximum for a given (fractional) .

Let us consider two cases, when and . We first begin with the former.

Case:

Again, let’s take a look at the pmf,

Similar to before, we are looking for the that maximizes the pmf for some arbitrary given . In the same way, we can ignore , since the choice of does not impact this value. Looking at the numerator, “”, since must be a positive fraction less than 1, then as increases it is strictly decreasing. In fact, for some given arbitrary , the largest value of the pmf is at , which evaluates to . All other values of start to decrease the overall value, since will itself be a fraction less than 1. In this case, when , the mode is , where is the floor of .

Case:

The last case is when is a fraction greater than 1. To understand, we can think about some example as follows (*Note: Let be the equivalent of* ):

At each step, for the next , the difference between it and is that the next equation is multiplied by . This term is always positive fraction but increases and then later decreases. When , then is greater than 1. When , then is less than 1. Therefore, up until , the output for will constantly be increasing by a factor larger than 1. However, at , the output will maximize, because at and beyond, it will continually multiple by a term that is less than 1. Therefore, the mode of the Poisson distribution when is a fraction is .

To summarize, if is an integer, then the mode is both and . If is a non-integer fraction, then the mode is .

* 1. Assume two categories and , equally probable *a priori*, distributed with Poisson distributions and . What is the Bayes classification decision?

Ans:

The Bayes decision rule is as follows,

Here, it is assumed that , therefore we only need to consider if . (*Note: Here, and are the likelihood functions rather than the probability mass functions.*)

The likelihood function for the Poisson distribution with parameter is as follows:

Therefore, the decision rule can be rewritten as follows,

where

* 1. What is the Bayes error rate?

Ans:

We choose to classify an example to if

Then let the following be shown:

*Note: since .*

Let , which is the ceiling of .

Here, Is the Bayes error rate.

1. Let for a two-category -dimensional problem with .
   1. Find , the minimum probability of error.

Ans:

We can choose to classify an observation to if

Then let the following be shown:

If , then the following holds:

So, we classify an example to if .

Let and .

* 1. Let and . Show that as the dimension approaches infinity. Assume that .

Ans:

From part a, we have that

Given that and , then can be rewritten as follows:

We have also seen before that the decision boundary is at , which in part b evaluates to,

Therefore, if an example falls in

then it is classified at , otherwise it is classified as and falls in

To show that as the dimension approaches infinity, let us refer back to . It has been divided into two parts, blue and red. First, we look at the blue section:

Next, let us look at the red section:

So, it follows that

where . Therefore, as the dimension approaches infinity.

1. Under the assumption that and , show that the general minimum risk discriminant function for a classifier with independent binary features is given by . What are and ?

Ans: (Reference: [5] pp.52-53)

Let , where are either 0 or 1 with probabilities

The class-conditional probabilities can be written as follows:

The likelihood ratio is as follows,

We are able to write out and in that format because of the assumption of conditional independence. This problem is a two-category case and so the classifier is known as a dichotomizer. Therefore, the separate discriminant functions can be combined into a single discriminant function. The one used here will be

Using this discriminant function leads to the following,

Since this discriminant function is linear in terms of , it can be written in the form of

where

and

**References**

[1] <https://john.cs.olemiss.edu/~ychen/courses/ENGR691F06/hw1/hw1sol.pdf>

[2] <https://www.statlect.com/probability-distributions/Poisson-distribution>

[3] <http://llc.stat.purdue.edu/2014/41600/notes/prob1804.pdf>

[4] <https://www.youtube.com/watch?v=lkjhwyrW8Io>

[5] Pattern Classification 2nd Ed. Duda, Hart, and Stork.

**Code Appendix**

library(latex2exp)  
### 1  
### a  
# sketch P(omega\_1 | x) = [p(x|omega\_1) \* P(omega\_1)] / p(x)  
cauchy <- function(x, a, b) {  
 (1 / (pi \* b)) \* (1 / (1 + ((x - a) / b)^2))  
}  
  
a1 <- 3; a2 <- 2; b <- 5; p\_omega1 <- 0.5; p\_omega2 <- 0.5  
p\_x <- function(x, p\_omega1, p\_omega2, a1, a2, b) {  
 p\_x\_given\_omega1 <- cauchy(x = x, a = a1, b = b)  
 p\_x\_given\_omega2 <- cauchy(x = x, a = a2, b = b)  
   
 prob\_x <- (p\_x\_given\_omega1 \* p\_omega1) +  
 (p\_x\_given\_omega2 \* p\_omega2)  
  
 return(prob\_x)  
}  
  
prob\_x <- p\_x(x = xs,  
 p\_omega1 = p\_omega1, p\_omega2 = p\_omega2,  
 a1 = a1, a2 = a2, b = b)  
  
omega\_given\_x <- function(x, p\_omega1, p\_omega2,  
 a1 = a1, a2 = a2, b = b) {  
 # Calculate p(x)  
 prob\_x <- p\_x(x = x,  
 p\_omega1 = p\_omega1, p\_omega2 = p\_omega2,  
 a1 = a1, a2 = a2, b = b)  
   
 # Calculate p(x|omega\_i)  
 p\_x\_given\_omega <- cauchy(x = x, a = a1, b = b)  
   
 cond\_prob <- (p\_x\_given\_omega \* p\_omega1) / prob\_x  
   
 return(cond\_prob)  
}  
  
xs <- seq(-3e2, 3e2, length.out = 1e3)  
ys <- omega\_given\_x(x = xs,  
 p\_omega1 = p\_omega1, p\_omega2 = p\_omega2,  
 a1 = a1, a2 = a2, b = b)  
  
plot(xs, ys, type = 'l',  
 main = TeX('$P(\\omega\_1 | x) \\; vs.\\; x$'),  
 xlab = TeX('$x$'), ylab = TeX('$P(\\omega\_1 | x)$'))  
### b  
prob\_error <- function(x) {  
 x <- abs(x)  
 0.5 - (1 / pi) \* atan(0.5 \* x)  
}  
  
xs <- seq(-5, 5, length.out = 1e3)  
plot(xs, prob\_error(xs), type = 'l',  
 main = TeX('$P(error) \\; vs. \\; | \\frac{a\_2 - a\_1}{b} | $'),  
 ylab = TeX('$P(error)$'), xlab = TeX('$ | \\frac{a\_2 - a\_1}{b} | $'))