Bayesian Decision Theory – 1

Introduction

We encounter lots of classification problems in real life. For example, an electronic store might need to know whether a particular customer based on a certain age, is going to buy a computer or not. Through this article, we are going to introduce a method named 'Bayesian Decision Theory' which helps us in making decisions on whether to select a class with 'x' probability or an opposite class with 'y' probability based on a certain feature.

Definition

Bayesian Decision Theory is a simple but fundamental approach to a variety of problems like pattern classification. The entire purpose of the Bayes Decision Theory is to help us select decisions that will cost us the least 'risk'. There is always some sort of risk attached to any decision we choose. We will be going through the risk involved in this classification later in this article.

Basic Decision

Let us take an example where an electronics store company wants to know whether a customer is going to buy a computer or not. So we have the following two buying classes:

w1 – Yes (Customer will buy a computer)

w2 – No (Customer will not buy a computer)

Now, we will look into the past records of our customer database. We will note down the number of customers buying computers and also the number of customers not buying a computer. Now, we will calculate the probabilities of customers buying a computer. Let it be P(w1). Similarly, the probability of customers not buying a customer is P(w2).

Now we will do a basic comparison for our future customers.

For a new customer,

If $P(w1) > P(w2)$, then the customer will buy a computer (w1)

And, if $P(w2) > P(w1)$, then the customer will not buy a computer (w2)

Here, we have solved our decision problem.

But, what is the problem with this basic Decision method? Well, most of you might have guessed right. Based on just previous records, it will always give the same decision for all future customers. This is illogical and absurd.

So we need something that will help us in making better decisions for future customers. We do that by introducing some features. Let's say we add a feature 'x' where 'x' denotes the age of the customer. Now with this added feature, we will be able to make better decisions.

To do this, we need to know what Bayes Theorem is.

Bayes Theorem and Decision Theory

For our class w1 and feature 'x', we have:

$P(w1 \mid x) = P(x \mid w1) * P(w1)P(x)$

There are 4 terms in this formula that we need to understand:

1. Prior – $P(w1)$ is the Prior Probability that w1 is true before the data is observed
2. Posterior – $P(w1 \mid x)$ is the Posterior Probability that w1 is true after the data is observed.
3. Evidence – $P(x)$ is the Total Probability of the Data
4. Likelihood – $P(x \mid w1)$ is the information about w1 provided by 'x'

$P(w1 \mid x)$ is read as Probability of w1 given x

More Precisely, it is the probability that a customer will buy a computer, given a specific customer's age.

Now, we are ready to make our decision:

For a new customer,

If $P(w1 | x) > P(w2 | x)$, then the customer will buy a computer (w1)

And, if $P(w2 | x) > P(w1 | x)$, then the customer will not buy a computer (w2)

This decision seems more logical and trustworthy since we have some features here to work upon and our decision is based on the features of our new customers and also past records and not just past records as in earlier cases.

Now, from the formula, you can see that for both our classes w1 and w2, our denominator P(x) is constant. So, we can utilize this idea and can form another form of decision as below:

If $P(x | w1)*P(w1) > P(x | w2)*P(w2)$, then the customer will buy a computer (w1)

And, if $P(x | w2)*P(w2) > P(x | w1)*P(w1)$, then the customer will not buy a computer (w2)

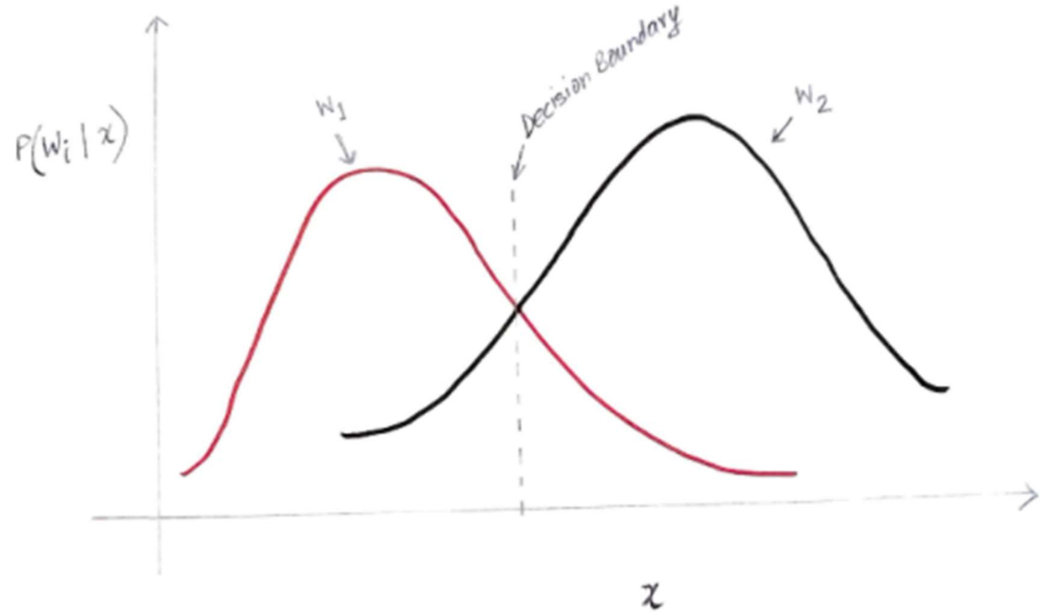
We can notice an interesting fact here. If somehow, our prior probabilities P(w1) and P(w2) are equal, we can still be able to make our decision based on our likelihood probabilities P(x | w1) and P(x | w2). Similarly, if our likelihood probabilities are equal, we can make decisions based on our prior probabilities P(w1) and P(w2).

Risk Calculation

As mentioned earlier, there is always going to be some amount of 'risk' or error made in the decision. So, we also need to determine the probability of error made in a decision. This is very simple and I will demonstrate that in terms of visualizations.

Let us consider we have some data and we have made a decision according to Bayesian Decision Theory.

We get a graph somewhat like below:



The y-axis is the posterior probability P(w(i) | x) and the x-axis is our feature ‘x’. The axis where the posterior probability for both the classes is equal, that axis is called our decision boundary.

So, at Decision Boundary:

P(w1 | x) = P(w2 | x)

So to the left of the decision boundary, we decide in favor of w1(buying a computer) and to the right of the decision boundary, we decide in favor of w2(not buying a computer).

But, as you can see in the graph, there is some non-zero magnitude of w2 to the left of the decision boundary. Also, there is some non-zero magnitude of w1 to the right of the decision boundary. This extension of another class over another class is what you call a risk or probability error.

Calculation of Probability Error

To calculate the probability of error for class w1, we need to find the probability that the class is w2 in the area that is to the left of the decision boundary.

Similarly, the probability of error for class w2 is the probability that the class is w1 in the area that is to the right of the decision boundary.

Mathematically speaking, the minimum error for class:

w1 is P(w2 | x)

And for class w2 is P(w1 | x)

You got your desired probability error. Simple, isn't it?

So what is the total error now?

Let us denote the probability of total error for a feature x to be P(E | x). Total error for a feature x would be the sum of all the probabilities of error for that feature x. Using simple integration, we can solve this and the result we get is:

P(E | x) = minimum (P(w1 | x) , P(w2 | x))

Therefore, our probability of total error is the minimum of the posterior probability for both the classes. We are taking the minimum of a class because ultimately we will give a decision based on the other class.

Conclusion

We have looked in detail at the discrete applications of Bayesian Decision Theory. You now know Bayes Theorem and its terms. You also know how to apply Bayes Theorem in making a decision. You have also learned how to determine the error in the decision you have made.