My Path Task | WorldQuant University

MODULE 3 | LESSON 1

BAYESIAN STATISTICS IN CLASSIFICATION

Reading Time	60 minutes
Prior Knowledge	Basic probability theory, Financial markets
Keywords	Bayesian statistics, Likelihood, Posterior, Prior

In the previous module, we looked at unsupervised machine learning. We will now use the explanatory variables to predict a target variable. This will be the first supervised machine learning algorithm that we'll cover.

1.The Bayesian Framework

We often want to find the probability that an observation will take on a certain class value, that is a success/failure or buy/hold, as an example of a binary classification problem. These events are known as a collective exhaustive set of events, in which there are only two possible outcomes, say and . Therefore, if we know the probability of one outcome, we can obtain the probability of the other by simply taking the complement, i.e., .

A probability of an event can depend on whether another event has taken place. For instance, we might consider the probability of a student passing an exam given that the student has studied for over 40 hours. In this case, the probability of passing the exam should be higher than if the student had studied for just four hours. If we denote the events:

- - The event that the student passes the exam
- - The event that the student prepares for the exam

We can write the statistical notation of as the probability of the student passing the exam given they have prepared for it. This is known as **conditional probability** and is the building block of Bayesian statistics.

The **joint probability** is the probability of events occurring together, i.e., is the probability of event *A* **and** *B* occurring. From the example above, would be the probability that the student prepares for the exam and passes, i.e., the joint probability of *A* and *S*.

1.1 Conditional Probability

For any two events and , the conditional probability rule is given by, (1.1).

Let's use the above rule in an example. A sample is taken of employees in the mining and finance sector who hold positions as either manager or director. Table 1.1 shows the sample in a contingency table. Based on the sample, suppose we wanted to know the probability of an employee being a manager given they are in the finance sector. Using (1.1), we have

Table 1

	Mining	Finance	TOTAL
Manager	580	700	1,280
Director	120	200	320
TOTAL	700	900	1,600

We now have an understanding of the basic probability theory needed for the Bayesian statistics approach. We will start by looking at Bayes's theorem and the Naive Bayes classifier that is based on the theorem.

Let us look at a video that will help us put this theory of probability into practice:

Access video transcript here

1.2 Bayes Theorem

Rearranging (1.1) we have the joint probability given by .

Keep in mind that the event of and are the same as and ; therefore, the above can also be written as (1.2).

Substituting (1.2) into (1.1), we obtain (1.3).

Equation (1.3) is known as Bayes theorem, which forms the basis for most algorithms using the Bayesian approach. Equation (1.3) can be extended to more than just one conditional variable. In other words, instead of just , we can have , , ... , . We write instead to keep it consistent with the notation of explanatory variables, and we will use instead of since the target or response variable is generally written as such. We will need this notation to explain the Naive Bayes classifier methodology. Now that we've seen the Bayes theorem, let's use this to build a classifier.

1.3 Naive Bayes

Suppose we have a target variable that takes on class values. With taking on a specific class value, say , equation (1.3) can be written as (1.4):

It is important to note that equation (1.4) is true provided the explanatory variables , , ... , are conditionally independent of , i.e. . This is the "naive" assumption made for the classifier and hence the name. Naive Bayes assumes conditional independence whereas Bayes's theorem does not. The denominator can also be written as,

by summing over all possible classes of the target variable. This is a property of marginalization. There are specific names given to the terms in (1.4), namely:

so we have (1.5):

To help you conceptualize this, please watch the following video.

Access video transcript here

1.4 Naive Bayes Example

Let us look at an example to understand the classifier. Table 1.2 contains credit risk data, specifically credit card defaulters and non-defaulters on their loans. There are two explanatory variables namely and , which are industry sector of employment and length of employment being more than 3 years or not. Determining the probability of defaulting on a loan is an important focus in the banking industry.

Table 2

	X_1	X_1	X_1	X_2	X_2	
	Mining	**Finance**	**Agriculture**	**Employed > 3 yrs**	**Employed 3 yrs or less**	**TOTAL**
Default (y=1)	250	600	150	300	700	1000
Non-default (y=0)	1250	6000	1750	1500	7500	9000
TOTAL	1500	6600	1900	1800	8200	10000

If we wanted the likelihood for an employee in the mining sector given they defaulted, we would use (1.1) to obtain (1.6):

and similarly, the likelihood of being employed for 3 years or less given the employee defaulted is

We could now determine the Naive Bayes probability of defaulting given the employee is in the mining sector and has been employed for three years or less using (1.4). For simplicity, let M = Mining and Eu = Employed 3 years or less.

Similarly, the . The classifier would assign the client as a non-defaulter since they have a higher probability of non-default. A useful reference for more details on Bayesian analysis can be found in chapter 12 of the Downey reading.

We have shown how the Naive Bayes classifier works and the assumptions made. The classifier can also be extended to continuous predictor variables with the assumption of normality. We've also included a link to another helpful video.

To consolidate the information presented above watch the following video:

Access video transcript here

2. Conclusion

In this lesson, we presented the methodology behind Bayesian statistics and the workings of the Naive Bayes classifier. We saw this probability approach used to address classification problems by using the prior and likelihood probabilities to calculate the chance of being classified as a particular class. A simplified example with few predictors is used to grasp the concept. The following lesson will implement this tool in a practical financial problem with more predictors. Therefore, we'll incorporate Python to assist with the heavy lifting.