Step by Step Introduction to Machine Learning. The Naïve Bayes Classifier.

Laith Adi
Laith_adi@hotmail.com
Saturday, May 15, 2021

Abstract

Regardless of who you are, where you live, or what you do for a living, everyone knows how vital artificial intelligence (AI) is to our world. Unlocking your phone using face ID, self-driving cars, your email automatically moving spam emails to the spam folder, Apples Siri, even a simple Google search, are some of the very few examples of how AI is implemented in our daily lives. Not to mention the role AI plays in Healthcare, Logistics, and other major sectors. With the significance of AI and considering that it is still a new topic with recent milestones such as autonomous vehicles and other deep learning applications, AI can cause a lot of confusion. This confusion can very much likely result in fear. “Robots are going to take over the world!” is a saying heard a lot and is usually said by people with knowledge on the subject drawn from sci-fi movies. In this article, I will be walking you through a concept of artificial intelligence called machine learning to enlighten those who are maybe confused or unsure. You do not need to have prior knowledge whatsoever to read this, as it is my job to make it as straight forward as possible, so anyone is able to understand. By the end of this article, you will have a better understanding of how AI works and your fear of robots taking over will vanish, for now at least.

Introduction

As I said earlier, we are going to focus on machine learning (ML), a concept of AI. Machine learning is teaching computers to solve problems autonomously. Using data, we can train a model (which is just a complex algorithm) to solve a problem. After training the model, the model can predict or conclude the target we are looking for, just like a human can do. Making it seem like it is acting like a human. This might be a bit abstract to wrap your head around, so let’s take the following example:

> You are studying for a math test. You have 100 textbook problems to use for practice, with the answers at the back of the textbook so you can check your solutions. You practice solving those 100 problems over and over, and then you write the test. During the test you are not given the solutions to the problems. After you write the test, the teacher marks your test by comparing your solution to the test’s solutions. Your teacher then gives you a mark, and based on that mark, you decide how you can improve for the next test.

In this example, the math test is the problem we are working towards. The 100 textbook questions are our data. Your approach to using the 100 questions to study for the test is our model. Now, to turn the example into a ML example, we use our data to train the model by feeding the questions to our model to learn. After training the model, we can then ask the model the test questions to solve the questions on its own. When the model attempts the test questions, we can calculate how well the model did by comparing the models’ solutions with the correct test solutions. With this score, we now can reflect and get a good idea on how to improve the model.

The math test example illustrates a supervised learning problem, one of the three machine learning problems. The three machine learning problems are:

(1) Supervised Learning
(2) Unsupervised Learning
(3) Reinforcement Learning

The three machine learning problems are the common approaches to solving any ML problem. For the sake of simplicity and relevance, we are going to focus only on supervised learning. In my opinion, this is the easiest to understand. You actually have a good idea of how it works from the earlier example. Supervised learning is when a model is trained using data that consists of two types of features:

(1) Descriptive Features: i.e., the textbook practice problems from our example.
(2) Target Features (or labels): i.e., the solutions to the textbook practice problems from our example.

Among each of the three machine learning problems, there are different models. Let's continue with our math test example to clearly explain what I mean by this.

> If your approach to studying is the model, then clearly there is more than one model since there is more than one way to study for a math test. Student "A" reviews their notes, then tries to solve all 100 practice problems, then checks the solutions to those problems. Student "B" jumps straight into the 100 practice problems, then checks their notes if they get stuck, then check the solutions to the problems. Student "C" looks straight at the solutions to the practice problems and copies them down. Student "A", "B", and "C" are all studying for the same test using the same 100 textbook problems and solutions, yet each of them used their own approach to studying. The different approaches can lead to different scores on the test. So, in this example, three different models that can be used to train using our data consisting of the 100 practice problems. Afterwards, we can test the models and compare scores to see which model works better and how to improve the model for the next test.

Now at this point, the goal is that you understand what machine learning is, how supervised machine learning problems work, and the idea of what ML models are and that there is more than one model in each of the three machine learning problems. We are now ready for the fun part! We are going to look at a model that uses probability-based learning, called The Naïve Bayes Classifier.

Bayes' Theorem

Before we get into the Naïve Bayes Classifier, we must understand what Bayes' Theorem is. To keep things simple, Bayes' Theorem is derived from conditional probability. Conditional probability is the probability that event "A" occurs based on the occurrence of event "B". We denote conditional probability as P(A|B), which can be read as "probability of A given B". Let's look at a very simple example.

ID	Toy	Happy
1	True	True
2	True	False
3	True	True
4	True	True
5	False	False

Table 1: Happy Children

Table above consists of three columns. **ID** column is just to identify the different children. **Toy** column, "True" if the child has a toy and "False" if the child does not have a toy. **Happy** column, "True" if the child is happy and "False" if the child is not happy. So, given the data in the table, what is the probability of a child having a toy?

P(**Toy**=True) = (**ID**: 1,2,3,4) / (**ID**: 1,2,3,4,5) = 4 / 5

There are 4 children with toys out of 5, so it makes sense to say that the probability of a child having a toy is 4/5. What about the probability of a child being not happy?

P(**Happy**=False) = (**ID**: 2,5) / (**ID**: 1,2,3,4,5) = 2 / 5

There are 2 children who are not happy out of 5 children in total. Thus, the probability of a child being not happy is 2/5. Now, what is the probability that a child has a toy given that they are happy?

P(**Toy**=True | **Happy**=True) = ???

This is a conditional probability. The condition is the child being happy. So, for all the children that are happy, what is the probability they have a toy? Well let's do this step by step. First, we observe which of the children is happy. Children with the **ID** 1,3,4 are happy since they have a value of "True" under the **Happy** column. Second, out of those three children, which of them have a toy? We can see under the **Toy** column that all three children have a value of "True", so they all have toys. We can go back to the question now and answer it.

P(**Toy**=True | **Happy**=True) = 3 / 3

The probability that a child has a toy given that they are happy is 3/3. Good job! Even though this is a very straight forward example, you have understood the concept of conditional probability.

Back to Bayes' Theorem, it solves a conditional probability using the following equation. Do not worry about how this equation is derived. It is beyond the scope of this article.

$$(\mathbf{1}) \qquad P(\boldsymbol{A}\,|\boldsymbol{B}) = \frac{P(\boldsymbol{B} \mid \boldsymbol{A}) \times P(\boldsymbol{A})}{P(\boldsymbol{B})}$$

<u>Generalization of Bayes' Theorem</u>

Wonderful, now you know what conditional probability is and how to solve for one. We have also stated the equation for the Bayes' Theorem. Up until now, we have only calculated the conditional probability of an event occurring based on the occurrence of one other event. What happens when a table has 10 columns, and we are trying to calculate the probability for one of those columns based on the other nine columns? In other words, what is the probability of event "A" occurring based on the occurrence of event "B", event "C", event "D", event "E", event "F", event "G", event "H", event "I", and event "J"? Or better yet, what if there are 100 columns or 1000 columns?

This might be the part where, as a beginner to this subject, might feel overwhelming. Well, I am here to tell you that you have nothing to feel overwhelmed about for two reasons. One, in this article we will be using simple examples to not confuse anyone. Two, in math, equations are often generalized so that they work for all scenarios of the problem the equation is intended for. Here is the generalized Bayes' Theorem for x number of columns.

$$(\mathbf{2}) \qquad P(\boldsymbol{Target} \mid \boldsymbol{col}[\mathbf{1}], \boldsymbol{col}[\mathbf{2}], \dots, \boldsymbol{col}[\boldsymbol{x}]) = \frac{P(\boldsymbol{col}[\mathbf{1}], \boldsymbol{col}[\mathbf{2}], \dots, \boldsymbol{col}[\boldsymbol{x}] \mid \boldsymbol{Target}) \times P(\boldsymbol{Target})}{P(\boldsymbol{col}[\mathbf{1}], \boldsymbol{col}[\mathbf{2}], \dots, \boldsymbol{col}[\boldsymbol{x}])}$$

The P(**Target**) is easy to calculate, but how do we calculate P(**col[1]**, **col[2]**, …, **col[x]** | **Target**) or P(**col[1]**, **col[2]**, …, **col[x]**)? Again, fear not because we can use the Chain Rule to solve those probabilities. The Chain Rule goes as follows:

$$(\mathbf{3})\qquad P(\boldsymbol{col}[\mathbf{1}], \boldsymbol{col}[\mathbf{2}], \dots, \boldsymbol{col}[\boldsymbol{x}]) = P(\boldsymbol{col}[\mathbf{1}]) \times P(\boldsymbol{col}[\mathbf{2}] \mid \boldsymbol{col}[\mathbf{1}]) \times \cdots \times P(\boldsymbol{col}[\boldsymbol{x}] \mid \boldsymbol{col}[\boldsymbol{x}-\mathbf{1}], \dots, \boldsymbol{col}[\mathbf{2}], \boldsymbol{col}[\mathbf{1}])$$

To apply the Chain Rule to conditional probability, it is the following:

$$(\mathbf{4})\qquad P(\boldsymbol{col}[\mathbf{1}], \boldsymbol{col}[\mathbf{2}], \dots, \boldsymbol{col}[\boldsymbol{x}] \mid \boldsymbol{Target}) = P(\boldsymbol{col}[\mathbf{1}] \mid \boldsymbol{Target}) \times P(\boldsymbol{col}[\mathbf{2}] \mid \boldsymbol{col}[\mathbf{1}], \boldsymbol{Target}) \times \cdots \times P(\boldsymbol{col}[\boldsymbol{x}] \mid \boldsymbol{col}[\boldsymbol{x}-\mathbf{1}], \dots, \boldsymbol{col}[\mathbf{1}], \boldsymbol{Target})$$

Conditional Independence

The final concept that must be introduced before we dive into Naïve Bayes Classifier is conditional independence. When two variables are independent, the event of either variable taking place has no effect on the probability of the other variable. Quick example, the probability that it will rain tomorrow does not affect the probability that I eat dinner. For conditional independence, two or more events may be independent if a separate event has happened.

For event "A" and event "B" that are conditionally independent based on the event "C", the probability of a joint event and conditional probability are:

$$(\mathbf{5})\qquad P(\boldsymbol{A} \mid \boldsymbol{B}, \boldsymbol{C}) = P(\boldsymbol{A} \mid \boldsymbol{C})$$

$$(\mathbf{6})\qquad P(\boldsymbol{A}, \boldsymbol{B} \mid \boldsymbol{C}) = P(\boldsymbol{A} \mid \boldsymbol{C}) \times P(\boldsymbol{B} \mid \boldsymbol{C})$$

Equation **(4)** can now be written as the following:

$$(\mathbf{7})\qquad P(\boldsymbol{col}[\mathbf{1}], \boldsymbol{col}[\mathbf{2}], \dots, \boldsymbol{col}[\boldsymbol{x}] \mid \boldsymbol{Target}) = P(\boldsymbol{col}[\mathbf{1}] \mid \boldsymbol{Target}) \times P(\boldsymbol{col}[\mathbf{2}] \mid \boldsymbol{Target}) \times \cdots \times P(\boldsymbol{col}[\boldsymbol{x}] \mid \boldsymbol{Target}) = \prod_{i=1}^{x} P(\boldsymbol{col}[\boldsymbol{i}] \mid \boldsymbol{Target})$$

Let's use equation **(7)** and substitute it into equation **(2)**.

$$(\mathbf{8})\qquad P(\boldsymbol{Target} \mid \boldsymbol{col}[\mathbf{1}], \dots, \boldsymbol{col}[\boldsymbol{x}]) = \frac{\left[\prod_{i=1}^{x} P(\boldsymbol{col}[\boldsymbol{i}] \mid \boldsymbol{Target})\right] \times P(\boldsymbol{Target})}{P(\boldsymbol{col}[\mathbf{1}], \dots, \boldsymbol{col}[\boldsymbol{x}])}$$

And just like that, we have the Bayes' Theorem with the assumption that there is conditional independence between the target variable and the conditions. We are very close to deriving the Naïve Bayes Classifier. Just two easy to understand changes must be applied.

One, the denominator of equation **(8)** applies a process called normalization. The Naïve Bayes Classifier does not apply normalization so we will simply remove it.

$$(\mathbf{9}) \qquad P(\boldsymbol{Target} \mid \boldsymbol{col}[\mathbf{1}], \dots, \boldsymbol{col}[\boldsymbol{x}]) = \left[\prod_{i=1}^{x} P(\boldsymbol{col}[\boldsymbol{i}] \mid \boldsymbol{Target})\right] \times P(\boldsymbol{Target})$$

Two, when predicting whether event happens or not, we need to calculate the probability that it does happen and the probability that it does not happen. Then we choose the higher probability as our conclusion. Quick example, say we want to predict whether or not it will rain tomorrow. Based on the information we have, the probability that it is going to rain is 0.34 and the probability it will not rain is 0.66. So, we calculate the probability of both outcomes then chose the higher of the two. So based of the probabilities, it will not rain tomorrow. Applying the same idea to any problem, and so tweaking equation **(9)** just a bit, we derived the Naïve Bayes Classifier.

$$(\mathbf{10}) \qquad P(\boldsymbol{q}) = \max_{l \,\in\, levels(\boldsymbol{Target})} \left[\prod_{i=1}^{x} P(\boldsymbol{q}[\boldsymbol{i}] \mid \boldsymbol{Target} = l)\right] \times P(\boldsymbol{Target} = l)$$

The Naïve Bayes Classifier

The previous section had a lot of equations, but you have made it. In this final section, we are going to run through an example, step by step. If you feel confused from the last section where we derived the Naïve Bayes Classifier, don't worry. Take a deep breath, it will make a lot of sense by the end of this section.

The Naïve Bayes Classifier, which we will refer to as NBC, is a supervised learning algorithm that uses probability-based learning. NBC uses probability-based learning because it is based on Bayes' Theorem, which is a probability theory and statistic. To explain NBC in the clearest way, we will work with the following data.

Patients ID	Fever	Dry Cough	Tiredness	Covid-19
1	True	True	True	True
2	False	True	False	False
3	False	False	False	False
4	True	True	True	True
5	True	False	True	False
6	True	True	False	True
7	False	True	True	False
8	True	False	True	True
9	True	True	False	True
10	False	True	True	True

Table 2: Covid-19 Patients

Let's talk about the above data to get a better grasp of it. The data consists of 5 columns. The **Patients ID** column consists of a unique ID for each patient. The rest of the columns consists of Boolean data type values. Meaning that the values can either be “True” or “False”. The **Fever** column tells us whether the patient has a fever or not. The **Dry Cough** column tells us whether a patient has a dry cough or not. **Tiredness** column tells us whether the patient is experiencing tiredness or not. The **Covid-19** column tells us if the patient has covid-19 or not. Pretty straight forward.

Since we know NBC is a supervised learning algorithm, we know that there are descriptive features and a target feature. Our target feature is what we are trying to train our model to predict. In this example, we want the model to predict the value for the **Covid-19** column, whether a patient has covid-19 or not. The descriptive features are **Fever**, **Dry Cough**, and **Tiredness** since they consist of useful information about the patients that we need to predict whether a patient has covid-19 or not. Notice how **Patients ID** is not a descriptive feature. This is because **Patients ID** has no effect on the chances that a patient will have covid-19 or not.

Here is how machine learning can be used to solve this problem. We will first clearly state some important information.

Problem statement	Whether or not a patient has covid-19.
Type of ML problem	Supervised learning.
Model	Naïve Bayes Classifier.
Data	Table 2: Covid-19 Patients.
Descriptive features	Fever. Dry Cough. Tiredness.
Target feature	Covid-19.

Restating NBC equation for convenience.

$$P(\boldsymbol{q}) = \max_{l \in levels(\boldsymbol{Target})} \left[\prod_{i=1}^{x} P(\boldsymbol{q}[\boldsymbol{i}] \mid \boldsymbol{Target} = l) \right] \times P(\boldsymbol{Target} = l) \qquad where,$$

$\boldsymbol{Target = Covid - 19}$

$l \in levels(\boldsymbol{Covid - 19}) = (\text{True, False})$

$x = 3\ (number\ of\ descriptive\ features)$

$\boldsymbol{q = Fever, Dry\ Cough, Tiredness}$

Great, now we have all the information we need. Let’s test our model to see if it works or not by using an existing patient in the Covid-19 Patients table and see if we can predict whether the patient has cavid-19. Let’s use patient with the ID number 2.

Patient ID	Fever	Dry Cough	Tiredness	Covid-19
2	False	True	False	False

If we were to apply this in code, then we would create our NBC model and feed the data we have from the table to our NBC model for it to train. Then enter the new patients’ symptoms in the trained model and observe the output of the model. But we are going to do this by hand to better illustrate how NBC works.

We will calculate the probability of the patient having covid-19 and the probability the patient not having covid-19. Then select the higher probability as our conclusion, or diagnostic. Since we are using a patient from the table and already know the result, we are expecting the probability to be higher that the patient does not have covid-19.

Let **Covid-19** = C19, **Fever** = **F**, **Dry Cough** = **DC**, and **Tiredness** = **T**, True = T, False = F.

$(*) \qquad P(\boldsymbol{C19} = \mathrm{T} \mid \boldsymbol{F} = \mathrm{F}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F})$
$(**) \qquad P(\boldsymbol{C19} = \mathrm{F} \mid \boldsymbol{F} = \mathrm{F}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F})$

Probabilities needed for NBC:

$$P(\boldsymbol{C19} = \mathrm{T}) = \frac{6}{10}$$

$$P(\boldsymbol{F} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{T}) = \frac{5}{6}$$

$$P(\boldsymbol{F} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{T}) = \frac{1}{6}$$

$$P(\boldsymbol{DC} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{T}) = \frac{5}{6}$$

$$P(\boldsymbol{DC} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{T}) = \frac{1}{6}$$

$$P(\boldsymbol{T} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{T}) = \frac{4}{6}$$

$$P(\boldsymbol{T} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{T}) = \frac{2}{6}$$

$$P(\boldsymbol{C19} = \mathrm{F}) = \frac{4}{10}$$

$$P(\boldsymbol{F} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{F}) = \frac{1}{4}$$

$$P(\boldsymbol{F} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{F}) = \frac{3}{4}$$

$$P(\boldsymbol{DC} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{F}) = \frac{2}{4}$$

$$P(\boldsymbol{DC} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{F}) = \frac{2}{4}$$

$$P(\boldsymbol{T} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{F}) = \frac{2}{4}$$

$$P(\boldsymbol{T} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{F}) = \frac{2}{4}$$

We will start with $(*)$ probability. The probability that the new patient has covid-19 given that they have a fever, no dry cough, and experiencing no tiredness. Using the equation **(10)**, we get the following:

$$P(\boldsymbol{C19} = \mathrm{T} \mid \boldsymbol{F} = \mathrm{F}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F})$$
$$= \left[\prod_{i=1}^{3} P(\boldsymbol{q}[\boldsymbol{i}] \mid \boldsymbol{C19} = \mathrm{T})\right] \times P(\boldsymbol{C19} = \mathrm{T})$$
$$= P(\boldsymbol{F} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{T}) \times P(\boldsymbol{DC} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{T}) \times P(\boldsymbol{T} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{T}) \times P(\boldsymbol{C19} = \mathrm{T})$$
$$= \left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right) \times \left(\frac{2}{6}\right) \times \left(\frac{6}{10}\right)$$
$$= \frac{60}{2160}$$
$$= \frac{1}{36}$$
$$\approx 0.02777$$

Using the same steps as above, we will now solve for $(**)$. The probability that the patient does not have covid-19 given that they have a fever, no dry cough, and experiencing no tiredness.

$$P(\boldsymbol{C19} = \mathrm{F} \mid \boldsymbol{F} = \mathrm{F}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F})$$

$$= \left[\prod_{i=1}^{3} P(\boldsymbol{q}[\boldsymbol{i}] \mid \boldsymbol{C19} = \mathrm{F})\right] \times P(\boldsymbol{C19} = \mathrm{F})$$

$$= P(\boldsymbol{F} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{F}) \times P(\boldsymbol{DC} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{F}) \times P(\boldsymbol{T} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{F}) \times P(\boldsymbol{C19} = \mathrm{F})$$

$$= \left(\frac{3}{4}\right) \times \left(\frac{2}{4}\right) \times \left(\frac{2}{4}\right) \times \left(\frac{4}{10}\right)$$

$$= \frac{48}{640}$$

$$= \frac{3}{40}$$

$$= 0.075$$

After solving for $(*)$ and $(**)$ we get the following probabilities.

$(*)$ $\quad P(\boldsymbol{C19} = \mathrm{T} \mid \boldsymbol{F} = \mathrm{F}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F}) \approx 0.02777$

$(**)$ $\quad P(\boldsymbol{C19} = \mathrm{F} \mid \boldsymbol{F} = \mathrm{F}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F}) = 0.075$

The probability that the patient does not have covid-19 based on no fever, dry cough, and no tiredness is greater than the probability that the patient does have covid-19. Thus, as expected, we can conclude that the **Covid-19** value for **Patient ID** 2 is False. If you go back and look at the Covid-19 Patients table, we can see that the value for **Covid-19** is False as well. Hurray, our model works and now we are ready to test the model using new data.

Imagine that we got hired at XYZ Hospital to implement our ML model to scan patient coming in in hopes to improve the efficiency of patient care during the pandemic. A new patient walks in complaining of symptoms that the nurse enters in the computer for our model to use. Here are the following symptoms for the new patient, with ID 18.

Patient ID	Fever	Dry Cough	Tiredness	Covid-19
18	True	True	False	???

The NBC model calculates the following probabilities as we did for $(*)$ and $(**)$. The probability that the new patient does have covid-19 given that the patient has a fever, dry cough, and no tiredness.

$$P(\boldsymbol{C19} = \mathrm{T} \mid \boldsymbol{F} = \mathrm{T}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F})$$

$$= \left[\prod_{i=1}^{3} P(\boldsymbol{q}[\boldsymbol{i}] \mid \boldsymbol{C19} = \mathrm{T})\right] \times P(\boldsymbol{C19} = \mathrm{T})$$

$$= P(\boldsymbol{F} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{T}) \times P(\boldsymbol{DC} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{T}) \times P(\boldsymbol{T} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{T}) \times P(\boldsymbol{C19} = \mathrm{T})$$

$= \left(\frac{5}{6}\right) \times \left(\frac{5}{6}\right) \times \left(\frac{2}{6}\right) \times \left(\frac{6}{10}\right)$

$= \frac{300}{2160}$

$= \frac{5}{36}$

≈ 0.13888

The probability that the new patient doesn't have covid-19 given that the patient has a fever, dry cough, and no tiredness.

$P(\boldsymbol{C19} = \mathrm{F} \mid \boldsymbol{F} = \mathrm{T}, \boldsymbol{DC} = \mathrm{T}, \boldsymbol{T} = \mathrm{F})$

$= \left[\prod_{i=1}^{3} P(\boldsymbol{q}[\boldsymbol{i}] \mid \boldsymbol{C19} = \mathrm{F})\right] \times P(\boldsymbol{C19} = \mathrm{F})$

$= P(\boldsymbol{F} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{F}) \times P(\boldsymbol{DC} = \mathrm{T} \mid \boldsymbol{C19} = \mathrm{F}) \times P(\boldsymbol{T} = \mathrm{F} \mid \boldsymbol{C19} = \mathrm{F}) \times P(\boldsymbol{C19} = \mathrm{F})$

$= \left(\frac{1}{4}\right) \times \left(\frac{2}{4}\right) \times \left(\frac{2}{4}\right) \times \left(\frac{4}{10}\right)$

$= \frac{16}{640}$

$= \frac{1}{40}$

$= 0.025$

Given the symptoms of the new patient, the probability that they have covid-19 is 0.13888. Which is greater than the probability that do not have covid-19, 0.025. Thus, the output value our NBC will predict is False for the **Covid-19** column for **Patient ID** 18. The nurse will see this and then proceed as she would for a patient that has covid-19. Not sure exactly how, I am not a doctor. Obviously.

Things to Keep in Mind

Since I wrote this with the intentions to provide an introduction for just about anyone to understand, I used straight forward examples and avoided as much terminology as possible. Also, since I skipped over small details and explanations, it might give off the wrong impression that ML is easy and that this is all you need to know. It is not. This does not even break the tip of the iceberg. This paper is meant to introduce one ML model thoroughly in simple terms through easy examples. Examples are critical as a simple explanation of ML does not do the topic justice. Without any examples, there will still be many questions left unanswered for those who are new to ML. I would like to give you guys something to think about. Things to investigate yourself and hopefully this will also show you how complicated things can get.

For one, what if our data consist of a column that holds values other than "True" and "False"? If we were to revisit the Covid-19 Patients table, but this time there is another column called **Temperature**. The **Temperature** column consists of decimal numbers like 60.34, 77.96, and so on. With float variables, that's what decimal number are called, NBC cannot calculate the probability simply by looking for temperatures that are the same. Every patient can have a different temperature. Hint: Probability density functions or binning is used. I will leave this topic up for you to discover on your own.

Another challenge that may arise when using NBC is what if a probability of 0 were to be calculated. For example:

$$P(\boldsymbol{A} \mid \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$$

$$= P(\boldsymbol{B} \mid \boldsymbol{A}) \times P(\boldsymbol{C} \mid \boldsymbol{A}) \times P(\boldsymbol{D} \mid \boldsymbol{A}) \times P(\boldsymbol{A})$$

$$= \frac{2}{4} \times \frac{0}{4} \times \frac{3}{4} \times \frac{4}{10}$$

$$= 0$$

Just because the final probability is 0, that does not mean it is accurate. It just so happens that one of the probabilities for an event happened to be 0. So, how would someone go about fixing this problem? Hint: To calculate the conditional property, the Laplacian smoothing is used. Another topic for you to delve deeper into on your own.

Conclusion

The best way to recap is to go back to the beginning. Machine learning is a concept within artificial intelligence. Using data, machine learning is training a machine to solve problems autonomously. Machine learning problems fall into three categories. First is supervised learning where the data is organised into descriptive features and target features. You can also say that the data is labelled. Second is unsupervised learning where in contrast to supervised learning, the data is not labelled. Lastly is reinforcement learning which focuses on the actions of an agent and the rewards gained in an environment. In this paper we discussed supervised learning. Unsupervised learning and reinforcement learning is up to you to explore. Within each of the ML problems, there are different models. Models are complex algorithms which can be based on probability learning just like the Naïve Bayes Classifier, or other learning styles like error-based learning, similarity-based learning, and more. Once a model is selected, engineers work on improving the model to perform better. The better the data is, among other factors, the more accurate a model is.

Today, machine learning is so powerful that it even outperforms humans in some tasks. Areas such as video games, image and object recognition, predictions, and more are all examples of where machine learning outperforms humans. This is not something to be frightened or sad about. It is something that should be encouraged and celebrated. Just take Tesla for example, everyone has heard of Tesla and yes, their cars are cool and all. Tesla reduces actual driving time. Drivers that might be distracted, under the influence, have slow reflexes, old, inexperienced, or simply any driver that just made a mistake, can all cause a serious accident. Tesla uses deep learning, a concept within machine learning, that lets the car drive on its own. Tesla is just one of many examples on why ML and AI is something that everyone should try and understand a little bit of.

So, with all that said, I will leave you with this. Artificial intelligence can change our whole world for the better. The potential is so big that we may not comprehend it right way. Artificial intelligence is a very big field, with many different concepts, machine learning being just one of them. It is our responsibility to push for a greater world. Artificial intelligence is one way to certainly do it. Find your own way and hold yourself accountable to it.