# 5 Naive Bayes

## Probability (Recap)

. The probability of an event is to compute how much likely this event is to happen.

- Famous example of the calibrated dice

| Event | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Probability | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

- The event in this case are the possible numbers that we can obtain while rolling the dice.
- Since the dice is calibrated, each of these numbers have a one out of six chance of happening or one of six probability of ocurring.

- Another famous example is Flipping the coin

| Event | Head | Tail |
|---|---|---|
| Probability | 1/2 | 1/2 |

- The events here are to obtain either Head or Tail
- The coins also calibrated, so we have equal probability for the 2 events

## Expectation of an event

. The expectation of a random variable is the summation of all possible values of a random variable, multiplied by the probability of each.

$$E(X) = \sum_{i=1}^{n} x_i \, p_i$$

$p_i$ ← Probability associated with each val

random variable ($X$)

↑ Values taken by the random variab ($x_i$)

Example :

| x | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| p | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

$E(X) = (1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$

$= 3.5$

Note: the expectation isn't the most probable value. It's however, similar to a mean value that a random variable can take

## The question is:

. What if we don't Know if the dice is calibrated?

. What if we don't Know if the coin is calibrated?

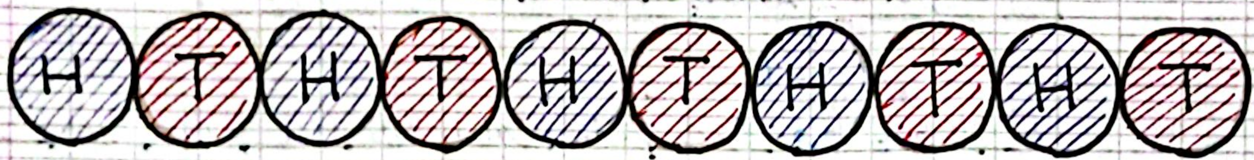
↳ The in this case we don't Know the probabilit and we need to approximate them.

This is how we can do that:

1. Let's take the example of the coin
   - We flip the coin for a large number of Times ($\infty$)
   - To approximate the probabilities, we calculate the probability of an event as follow:

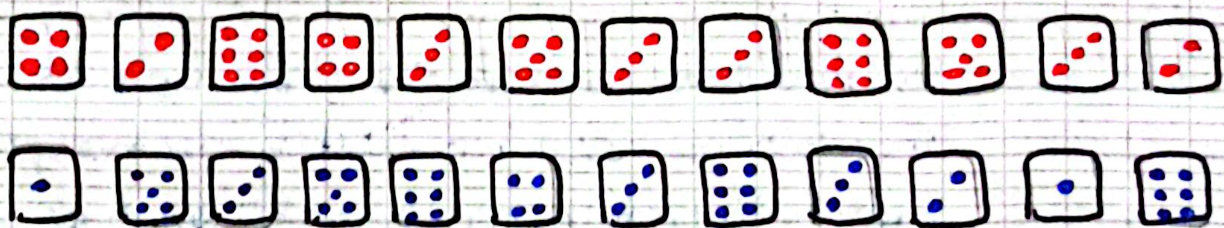
$$P(\text{event}) = \frac{\text{Number of occurences of that event}}{\text{Total number of trials}}$$

- So let's say we flip the coin 10 times (in real life it should be a lot more than that
- We obtain the following results




$$\left.\begin{aligned} P(\text{Head}) &= 5/10 = 0.5 \\ P(\text{Tail}) &= 5/10 = 0.5 \end{aligned}\right\} \begin{aligned} &\text{Equal Probabilities} \\ &\text{Coin is Calibrated} \end{aligned}$$

2. Let's take another example, This time we are rolling 2 dices for 12 Times



- What's the probability that dice 2 gives 6 given that the dice 1 gives 3

. So here we have 2 events, one that's given (dice 1 gives 3), so it happened and one that could happen as a result (dice 2 gives 6)
. This is what we call the conditional probability

$$P(\text{event } A / \text{event } B) = \frac{P(\text{event } A \text{ and event } B)}{P(\text{event } B)}$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

. To answer the question:

$$P(\text{dice 2 gives 6} / \text{dice 1 gives 3}) = \frac{P(6 \text{ and } 3)}{P(3)}$$

→ . There are 3 cases where we obtained 6
. There are 2 cases where we obtained 6 and 3
. There are 4 cases where we obtained 3
. There are in total 12 cases

$$\begin{cases} P(6 \text{ and } 3) = 2/12 \\ P(3) = 4/12 \end{cases} \Rightarrow P(2 \text{ gives } 6 / 1 \text{ gives } 3) = 2/4$$

→ Naive Bayes
. Supervised ML algorithm that uses Bayes theorem in order to carry out classification problems
. Bayes theorem relies on the concept of conditional probability

## Baye's Theorem

- A and B are 2 events
- $P(A/B)$ is the probability of A given B is True
- $P(B/A)$ is the probability of B given A is True
- $P(A)$ and $P(B)$ are the independent probabilities of A and B respectively

$$P(A/B) = P(A \cap B) / P(B) \quad , \quad P(A \cap B) = P(B \cap A)$$

$$P(B/A) = P(B \cap A) / P(A) \qquad P(B \cap A) = P(B/A).P(A)$$

$$\Rightarrow P(A/B) = \frac{P(B/A).P(A)}{P(B)}$$

- How does this apply to ML?
- In ML, A is the class that we want to predict (Y), and B is the given set of features (X)

→ The equation becomes as follow:

$$P(Class / Features) = \frac{P(Features/Class).P(Class)}{P(Features)}$$

From this equation, we know that we can predict the class of an example given a set of features that we have. And all this by relying on the labeled data that we have that can tell us the probability of having specific features, given that we are in a specific class.

## 2. Naive Bayes - Example

- To understand how Naive Bayes functions, we will consider an example with one feature
- The question we want to answer: Kids will play if the weather is Sunny?

| Weather | Play |
|---|---|
| Sunny | No |
| Overcast | Yes |
| Rainy | Yes |
| Sunny | Yes |
| Sunny | Yes |
| Overcast | Yes |
| Rainy | No |
| Rainy | No |
| Sunny | Yes |
| Rainy | Yes |
| Sunny | No |
| Overcast | Yes |
| Overcast | Yes |
| Rainy | No |

**Step 1: Convert the data set into a frequency table**

- In order to build our frequency table, we count the number of occurances of each event given the conditions (Sunny, overcast, rainy)

| Frequency Table | | |
|---|---|---|
| Weather | No | Yes |
| Overcast | 0 | 4 |
| Rainy | 3 | 2 |
| Sunny | 2 | 3 |
| Total | 5 | 9 |

**Step 2: Create likelihood table**

In order to calculate the likelihood of Kids playing with respect to the weather condition, we begin by computing the probability of each condition and event.

↳ Ex:

$$P(\text{Rainy}) = \frac{\text{Nb of occurances of rainy days}}{\text{Total Nb of Occurences}} = 5/14$$

| Likelihood Table | | | | |
|---|---|---|---|---|
| Weather | No | Yes | | |
| Overcast | 0 | 4 | =4/14 | 0.29 |
| Rainy | 3 | 2 | =5/14 | 0.36 |
| Sunny | 2 | 3 | =5/14 | 0.36 |
| All | 5 | 9 | | |
| | =5/14 | =9/14 | | |
| | 0.36 | 0.64 | | |

→ These probabilities will help us with the naive bayes formula.

. We use the Bayes equation to get the probability for each class given a condition

$$P(Yes/Sunny) = \frac{\overset{3/9}{P(Sunny/Yes)} \cdot \overset{9/14}{P(Yes)}}{\underset{5/14}{P(Sunny)}}$$

$= 0.6$
(label. True)

⇒ This prove that the players will play if the weather is sunny, using the bayes theorem, which basically is the chances are more than half

As A and B are independent
⇒ $p(A \cap B) = p(A) \cdot p(B)$
so $p(A \text{ and } B / C)$
$= p(A \cap B / C)$
$= p(A/C) \cdot p(B/C)$

↳ What if we have several features?

. One of the characteristics of Naive Bayes is that it considers the features to be independent

. Recall: If X and Y are 2 independent s:

↳ $P(X \cap Y) = P(X) \cdot P(Y)$

$$\Rightarrow \begin{cases} P(X/Y) = P(X) \\ P(Y/X) = P(Y) \end{cases}$$

↳ $P(X \cap Y) = P(X) \cdot P(Y)$

$$\Rightarrow P(\underbrace{X_1 \overset{and}{,} X_2}_{\text{ind. features}} / \underbrace{Y}_{\text{class}}) = P(X_1/Y) \cdot P(X_2/Y)$$

. But how we can use this information?

. To understand, let's assume we have 2 features $F_1$ and $F_2$. Bayes rule will become as follows:

$$P(Class / F_1, F_2) = \frac{P(F_1, F_2 / Class) . P(class)}{P(F_1 \cap F_2)}$$

$$\Rightarrow P(Class / F_1, F_2) = \frac{P(F_1 / Class) . P(F_2 / Class) . P(Class)}{P(F_1) . P(F_2)}$$

. The same idea applies if we had more features

. So in case of 2 features and 2 classes, for example, we compute the following probabilities:

$$P(Class=1 / F_1, F_2) = \frac{P(F_1 / Class=1) . P(F_2 / Class=1) . P(Class=1)}{P(F_1) . P(F_2)}$$

$$P(Class=0 / F_1, F_2) = \frac{P(F_1 / class=0) . P(F_2 / Class=0) . P(class=0)}{P(F_1) . P(F_2)}$$

. Note that the values that we'll get will not represent the real probability unless the features we're taking are actually really independent

. As we can see, the denominator in the 2 formulas look the same, in respect to different classes } So we can simplify and remove it.

$$P(Class / F_1, F_2) \underset{\text{proportional}}{\propto} P(F_1 / Class) . P(F_2 / Class) . P(class)$$

. After simplifying, we can see that the probability of a class given that we have $F_1$ and $F_2$ is proportional to the numerator that we have earlier.

. The predicted class is the one that has the highest probability given the two features $F_1$ and $F_2$

## → Naive Bayes Applications

. 1) Sentiment Analysis

It enables companies to understand what their customers like and dislike, for example.

2) Gene Analysis

Naive Bayes has been used to analyse genes. It has been mainly based on already existing genes available in databases, enabling us to classify certain types of genes and genetic mutations.

## → Advantage and Disadvantages of Naive Bayes

| Advantages | Disadvantages |
| --- | --- |
| . Fast to train | . Assumes the conditional independence of features (loss in accuracy) |
| . Insensitive to irrelevant features | . Dependencies among features exist but cannot be modeled by Naive Bayes. |
| . Can handle real and discrete data | |