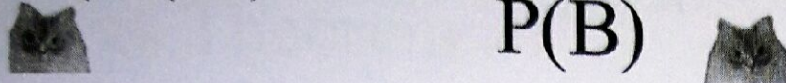


Supervised Learning and Naive Bayes Classification

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$


Part one describes how naive bayes classifier works. **Part two** consist of a programming exercise in Python using sklearn library that provides Naive Bayes Classifiers.

Imagine two people Alice and Bob whose word usage pattern you know. To keep example simple, lets assume that Alice uses combination of three words [love, great, wonderful] more often and Bob uses words [dog, ball wonderful] often.

Lets assume you received an anonymous email whose sender can be either Alice or Bob. Lets say the content of email is *"I love beach sand. Additionally the sunset at beach offers wonderful view"*

Can you guess who the sender might be?

Well if you guessed it to be Alice you are correct. Perhaps your reasoning would be the content has words love, great and wonderful that are used by Alice.

Now let's add a combination and probability in the data we have. Suppose Alice and Bob uses following words with probabilities as show below. Now, can you guess who is the sender for the content : *"Wonderful Love."*

Alice	Bob
Love [0.1]	Wonderful [0.5]
Wonderful [0.1]	Love [0.3]
Great [0.8]	Deal [0.3]

Now what do you think?

If you guessed it to be Bob, you are correct. If you know mathematics behind it, good for you. If not, don't worry we shall do it in next section. This is where we apply Bayes Theorem.

Bayes Theorem

$$\underset{\text{cat}}{P(A | B)} = \frac{P(B | A)P(A)}{P(B)} \underset{\text{cat}}{}$$

It tells us how often A happens *given that B happens*, written $P(A|B)$, when we know how often B happens *given that A happens*, written $P(B|A)$, and how likely A and B are on their own.

- $P(A|B)$ is “Probability of A given B”, the probability of A given that B happens
- $P(A)$ is Probability of A
- $P(B|A)$ is “Probability of B given A”, the probability of B given that A happens
- $P(B)$ is Probability of B

When $P(\text{Fire})$ means how often there is fire, and $P(\text{Smoke})$ means how often we see smoke, then:

$P(\text{Fire}|\text{Smoke})$ means how often there is fire when we see smoke.
 $P(\text{Smoke}|\text{Fire})$ means how often we see smoke when there is fire.

So the formula kind of tells us “forwards” when we know “backwards” (or vice versa)

Example: If dangerous fires are rare (1%) but smoke is fairly common (10%) due to factories, and 90% of dangerous fires make smoke then:

$$P(\text{Fire}|\text{Smoke}) = P(\text{Fire}) P(\text{Smoke}|\text{Fire}) = 1\% \times 90\% = 9\% P(\text{Smoke}) 10\%$$

In this case 9% of the time expect smoke to mean a dangerous fire.

Now can you apply this to our Alice and Bob example?

Naive Bayes Classifier

Naive Bayes classifier calculates the probabilities for every factor (here in case of email example would be Alice and Bob for given input feature). Then it selects the outcome with highest probability.

This classifier assumes the features (in this case we had words as input) are independent. Hence the word naive. Even with this it is powerful algorithm used for

- **Real time Prediction**
- **Text classification/ Spam Filtering**
- **Recommendation System**

So mathematically we can write as,

If we have a certain event E and test actors x_1, x_2, x_3 , etc.

We first calculate $P(x_1 | E)$, $P(x_2 | E)$... [read as probability of x_1 given event E happened] and then select the test actor x with maximum probability value.