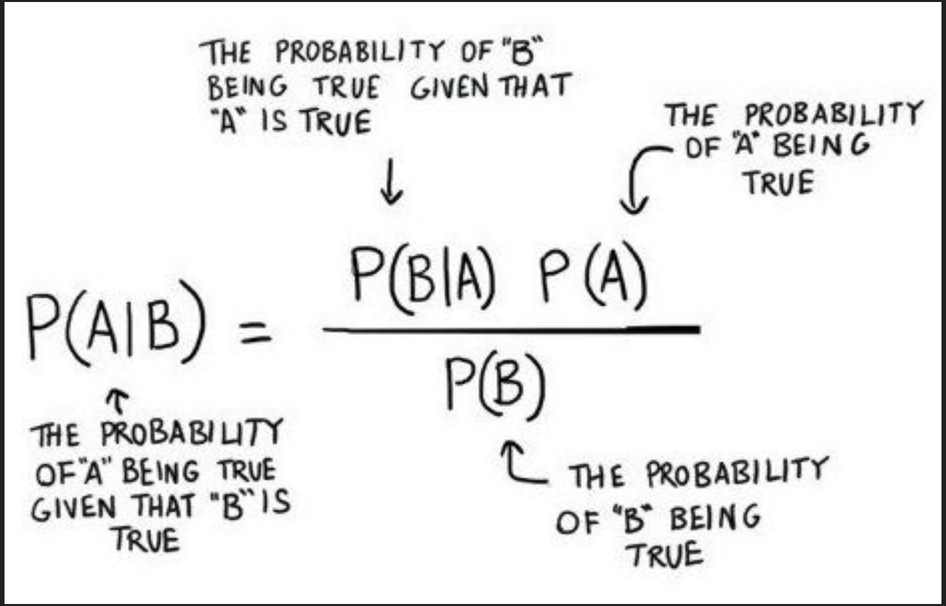
Naive Bayes Classifier | ML LAB 6 | VTU

By Prajwal Mani

What is Bayes Theorem?



Bayes theorem talks about Conditional probability.

Multinomial Naive Bayes theorem

$$V_{NB} = \underset{vj \in v}{\operatorname{argmax}} P(v_j) \prod_{i \in positions} P(a_i|v_j)$$

Why choose Multinomial Naive Bayes?

- Because our dataset has discrete values.
- Eg. movie ratings ranging 1 and 5 as each rating will have certain frequency to represent)
- This works well for data which can easily be turned into counts, such as word counts in text.
- It is regularly used in natural lanaguage processing (NLP) problems

What is CountVectorizer?

CountVectorizer is a great tool provided by the scikit-learn library in Python. It is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text.

Example of CountVectorizer

document = [“One Geek helps Two Geeks”, “Two Geeks help Four Geeks”, “Each Geek helps many other Geeks at GeeksforGeeks.”]

	at	each	four	geek	geeks	geeksforgeeks	help	helps	many	one	other	two
document[0]	0	0	0	1	1	0	0	1	0	0	0	1
document[1]	0	0	1	0	2	0	1	0	0	0	0	1
document[2]	1	1	0	1	1	1	0	1	1	0	1	0

Accuracy Score

In multilabel classification, the function returns the subset accuracy. If the entire set of predicted labels for a sample strictly match with the true set of labels, then the subset accuracy is 1.0; otherwise it is 0.0.

Precision Score

- The precision is the ratio **tp / (tp + fp)** where **tp** is the number of **true positives** and **fp** the number of **false positives**. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.
- The best value is 1 and the worst value is 0.

Recall Score

- The recall is the ratio **tp / (tp + fn)** where **tp** is the number of **true positives** and **fn** the number of **false negatives**. The recall is intuitively the ability of the classifier to find all the positive samples.
- The best value is 1 and the worst value is 0.

Confusion Matrix

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)