Naive Bayes Classifiers

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem.
Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred.
Bayes' theorem is stated mathematically as the following equation:

P(A|B) = P(B|A) P(A) / P(B)

- The formula is trying to find probability of event A, given the event B is true. Event B is also termed as evidence.
- P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).
- P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.

Assumptions:
The fundamental Naive Bayes assumption is that each feature makes an:

- Independent (We assume that no pair of features are dependent).
- Equal (each feature is given the same weight or importance)

contribution to the outcome.

Naive Bayes Classifiers

Example:
(source: https://www.geeksforgeeks.org/naive-bayes-classifiers/)

We want to see what the probability of playing golf or not on a day given certain weather conditions.

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

To see if you would play if the weather was Outlook "Sunny", Temperature "High", Humidity "Normal", Wind "False" you work out the probability of playing and not playing based on the probabilities of previous outcomes in the table above.

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

Naive Bayes Classifiers

$$P(No|today) = \frac{P(SunnyOutlook|No)P(HotTemperature|No)P(NormalHumidity|No)P(NoWind|No)P(No)}{P(today)}$$

$$P(Yes|today) \propto \frac{2}{9} \cdot \frac{2}{9} \cdot \frac{6}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} \approx 0.0141$$

$$P(No|today) \propto \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{5}{14} \approx 0.0068$$

These numbers can be converted into a probability by making the sum equal to 1 (normalization):

$$P(Yes|today) = \frac{0.0141}{0.0141+0.0068} = 0.67$$

$$P(No|today) = \frac{0.0068}{0.0141+0.0068} = 0.33$$

Since the probability of playing is greater than not playing then the prediction is that golf would be played.

Pros and Cons
Pros:

- It is easy and fast to predict class of test data set. It also perform well in multi class prediction
- When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.
- It performs well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).

Cons:

- If categorical variable has a category (in test data set), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction.
- On the other side naive Bayes is also known as a bad estimator, so the probability outputs from predict_proba are not to be taken too seriously.
- Another limitation of Naive Bayes is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

Naive Bayes Classifiers

4 Applications of Naive Bayes Algorithms

1. Real time Prediction
2. Multi class Prediction
3. Text classification/ Spam Filtering/ Sentiment Analysis
4. Recommendation