Naive Bayes -

- K-NN is neighborhood classfication technique.
- Naive Bayes is Probability based classification technique

Conditional Probability-

- Calculate the probability of the first event A given that second event B has already happened.

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

Independent Event- A, B are said to be independent if

$$P(A/B) = P(A)$$
$$P(B/A) = P(B)$$

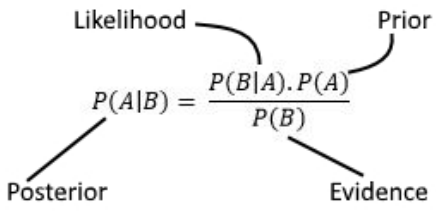
- Eg A: Getting value of 6 in die 1 throw D1=6.
  B: Getting a value of 3 in dies 2 thrown, D2 =3, Here both are independent event, impact of D1 does not affect D2.

Mutually Exclusive-

- Mutually exclusive events are those events where two events cannot happen together.
- Eg. Toss of a coin. Getting a head and a tail are mutually exclusive because we can either get heads or tails but never both at the same in a single coin toss.
- If P(A/B)=P(B/A)=0, then A & B are said to Mutually Exclusive.

Bayes Theorem -

Bayes Theorem finds the probability of an event occurring given the probability of another event that has already occurred.

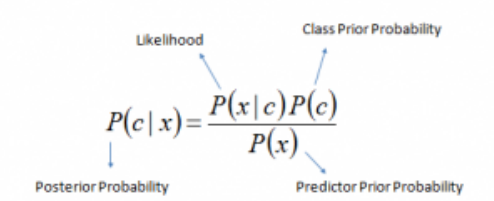


- Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.
- P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).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

Naive Bayes

It is a classification technique based on Bayes Theorem, with the assumption that the **presence of a particular feature in a class is independent to the presence of any other feature.**

- Useful for very large data sets
- For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as ‘Naive’.



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

What are the Pros and Cons of Naive Bayes?

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 perform 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. This is often known as “Zero Frequency”. To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.
- 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.

Applications of Naive Bayes Algorithms

- **Real time Prediction-** Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.
- **Multi class Prediction** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
- **Text classification/ Spam Filtering/ Sentiment Analysis-** Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in

Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)
- **Recommendation System-** Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not

Laplace Smoothing or Additive Smoothing(Its not Laplacian Smoothing)

- Laplace Smoothing is a technique to smooth categorical data.
- Laplace Smoothing is introduced to solve the problem of zero probability.
- Here we add alpha to numerator and alpha*k to denominator, usally k=2, alpha=

Imbalance Dataset in Naive Bayes

1. Up-sampling (Over sampling) the minority class
2. Down-sampling (Under sampling) the majority class

Outliers

To Handle Outlier-

1. If a word occur fewer than 10 times, then just ignore that word.
2. Laplace Smoothing

Missing Value

1. Text-data - No case of missing data
2. Categorical features- Consider NaN as a new Category.
3. Numerical Feature - We use standard feature imputation like mean, median, K-NN and we can use Gaussian Naive Bayes.

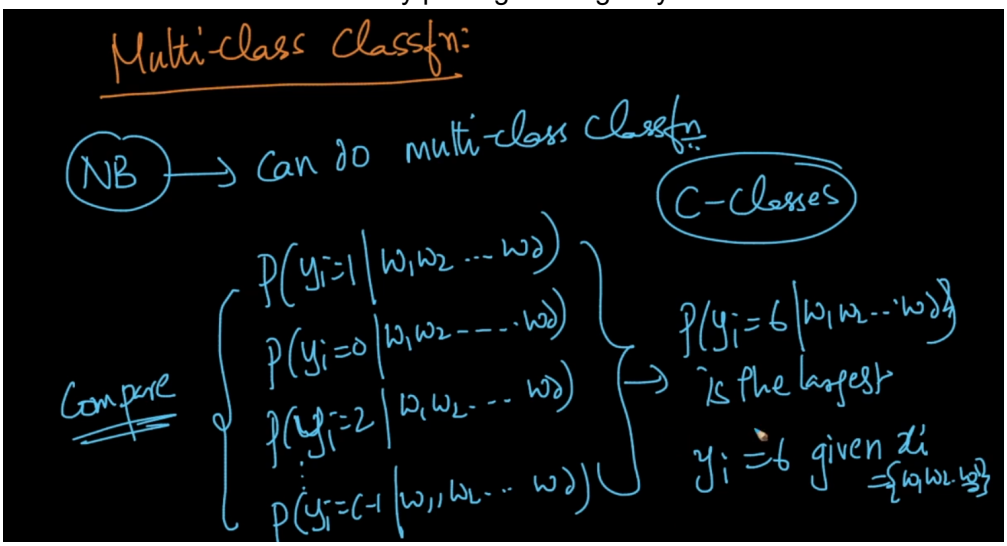
Gaussian Naive Bayes

Refer- https://machinelearningmastery.com/naive-bayes-for-machine-learning/ (https://machinelearningmastery.com/naive-bayes-for-machine-learning/)

- Assume its follow Conditional independent, which means each of feature is independent of other given the class label.
- Naive Bayes used for real-valued attributes, most commonly by assuming a Gaussian distribution.
    - Its used to estimate the distribution of the data, but the Gaussian (or Normal distribution) is the easiest to work with because you only need to estimate the mean and the standard deviation from your training data.

Ques- Can NB do Multiclass Classification?

Ans- NB can do multiclass classification by picking the largest yi.



Ques: Can NB use Distance or Similarity matrix?

Ans- No it can not

Ques- Can NB use with Large Dimensionality?

Ans- Yes, it can be used, to maintian numberical stability we use with log-probability.

Best and Worst Case of Naive Bayes

1. Conditional Independence of Feature-
    - If its True- NB perform very well.
    - If its False- NB degrade.
    - If some feature have dependent even then NB works well.

2. Text-Classfication
    - Used in detecting Email Spam, Detect Review Polarity(+ve, -ve)

3. Categorical Feature
    - NB is used when we have Categorical feature, Real-valued feature.

4. Interpretable, feature-importance-
    - Runtime complexity- low
    - Train-time complexity- low
    - Run-time space complexity- low

5. Easily Overfit if we dont use Laplace Smoothing(Worst Case)

Refer -
Naive Bayes - https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/ (https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/)