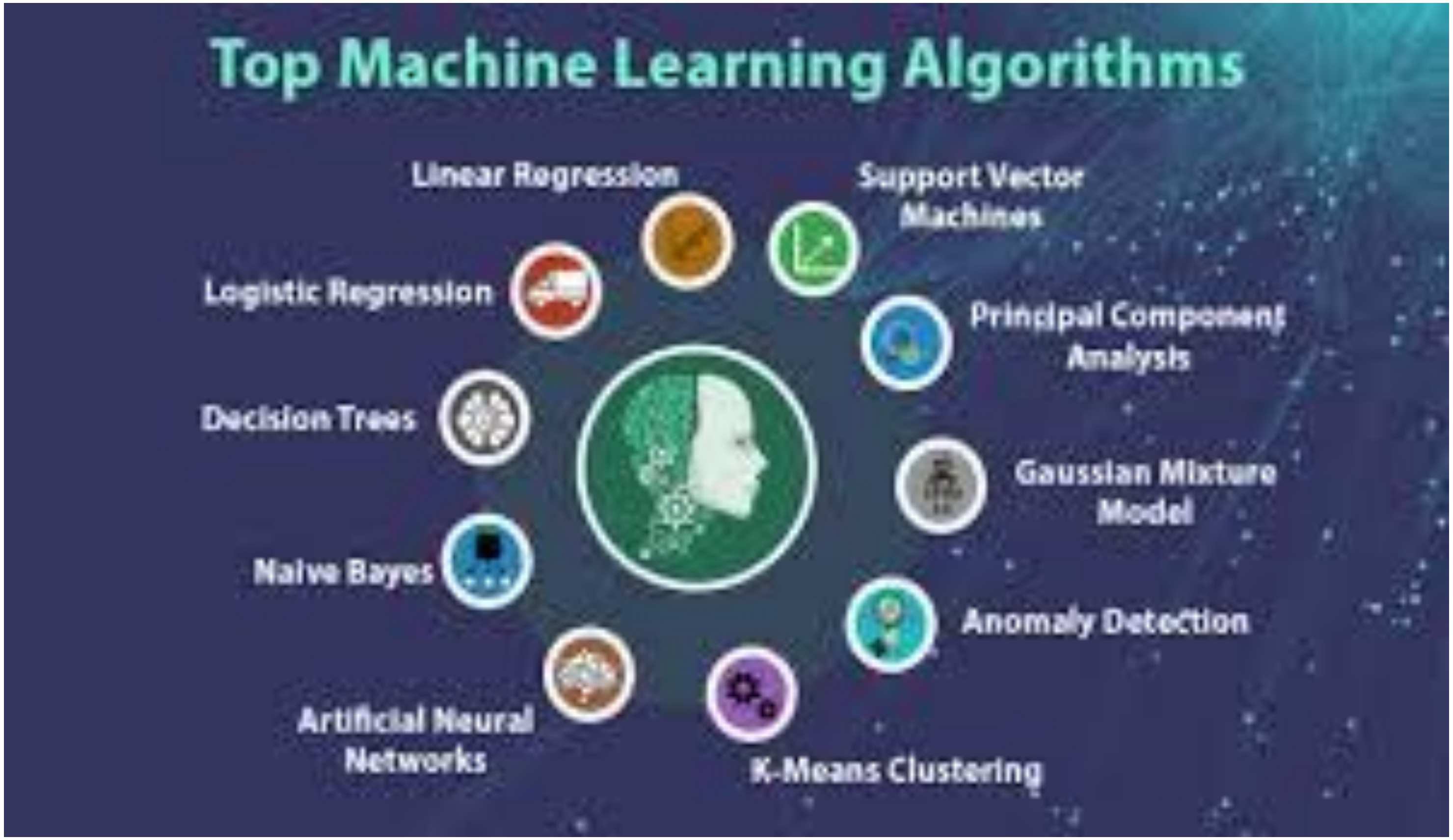
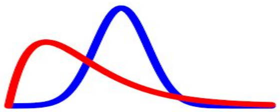


Key ML Algorithms

Agenda

- Naive Bayes Algorithm
- Multinomial Naive Bayes
- Support Vector Regression
- R-Squared Intuition
- Adjusted R-Squared Intuition

Naive Bayes



- **What is a classifier?**

  A classifier is a machine learning model that is used to discriminate different objects based on certain features.

- **Principle of Naive Bayes Classifier:-**

  A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem.

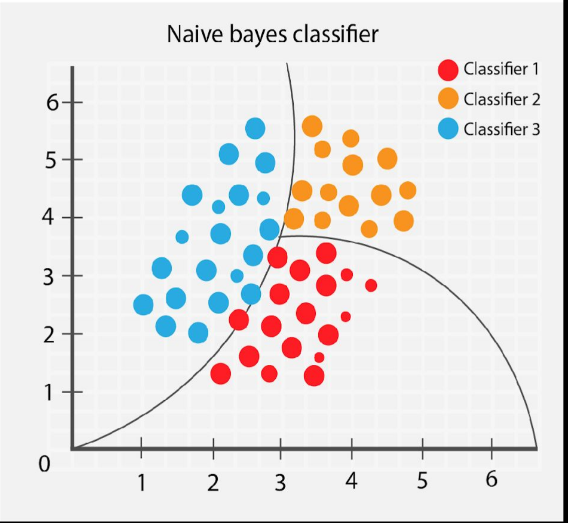
Bayes Theorem

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

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

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$



- Using Bayes theorem, we can find the probability of **A** happening, given that **B** has occurred.
- Here, **B** is the evidence and **A** is the hypothesis. The assumption made here is that the predictors/features are independent.
- That is presence of one particular feature does not affect the other. Hence it is called naive.

Conditional Probability:

- In the domain of probability theory, conditional probability is the measure of an event A occurring when another event say B has taken place. This is represented by and is read as "the conditional probability of A given B"…(A|B)
- One of the example in the domain of conditional probability is flipping the coin…the chances of having a head or tail are equal. IN other words, the probability of any of the events is 0.5. So, conditional probability talks about the chances of having a head once we already had a tail. Bayes theorem provides a mathematical model for calculating these.

Bayes Theorem:

- It describes the probability of occurrence of an event, based upon the existing knowledge about the conditions that are related to that specific event.
- Say, diabetes happens at some particular age X, and then by using the Bayes theorem, the age of a person can be used to forecast the chances that they will have diabetes and the results will be much better as compared to a situation when we had no idea about their age.

Example- Picnic Day

You are planning a picnic today, but the morning is cloudy

- Oh no! 50% of all rainy days start off cloudy!
- But cloudy mornings are common (about 40% of days start cloudy)
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of the rain?

- We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.
- The chance of Rain given Cloud is written P(Rain|Cloud)
- So let's put that in the formula:

$$P(\text{Rain}|\text{Cloud}) = \frac{P(\text{Rain})\,P(\text{Cloud}|\text{Rain})}{P(\text{Cloud})}$$

- P(Rain) is Probability of Rain = 10%
- P(Cloud|Rain) is Probability of Cloud, given that Rain happens = 50%
- P(Cloud) is Probability of Cloud = 40%

$$P(\text{Rain}|\text{Cloud}) = \frac{0.1 \times 0.5}{0.4} = .125$$

Or a 12.5% chance of rain. Not too bad, let's have a picnic!

INTERNSHIPSTUDIO

How Naive Bayes algorithm works?

We use Weather and play golf dataset here

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

INTERNSHIPSTUDIO

Frequency table for each attribute

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$$

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(Yes) = 9/14 = 0.64$$

Posterior Probability: $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
		9	5	14

$$P(x \mid c) = P(Sunny \mid No) = 2/5 = 0.4$$

$$P(x) = P(Sunny) = 5/14 = 0.36$$

$$P(c) = P(No) = 5/14 = 0.36$$

Posterior Probability: $P(c \mid x) = P(No \mid Sunny) = 0.40 \times 0.36 \div 0.36 = 0.40$

The likelihood tables for all four predictors

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

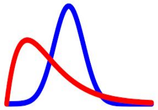
		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

The class with the highest posterior probability is the outcome of prediction.

Naïve Bayes Classifier – Pros & Cons

- Advantages
  - Easy to implement
  - Good results obtained in most of the cases
- Disadvantages
  - Assumption: class conditional independence, therefore loss of accuracy
  - Practically, dependencies exist among variables
    - E.g., hospitals: patients: Profile: age, family history, etc.
      Symptoms: fever, cough etc., Disease: lung cancer, diabetes, etc.
    - Dependencies among these cannot be modeled by Naïve Bayes Classifier
- How to deal with these dependencies? Bayesian Belief Networks

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).

What are the Pros and Cons of Naive Bayes?

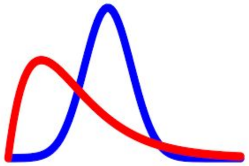
Cons:

- 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 is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

Types of Naive Bayes

Scikit learn (python library) will help here to build a Naive Bayes model in Python. There are three types of model under the scikit-learn library :-

- Gaussian: It is used in classification and it assumes that features follow a normal distribution.
- Multinomial: It is used for discrete counts. For example, a text classification problem.
- Bernoulli: The binomial model is useful if your feature vectors are binary (i.e. zeros and ones).



Q.1 Define classifier in terms of ML?

Q.2 What is Bayes Theorem. Please write its formula?

Q.3 Write the Probability for picnic example in Bayes theorem?

Q.4 Explain how Naïve Bayes Classifier works?

Q.5 What are the advantages and disadvantages of Naïve Bayes Classifier ?

Q.6 Explain in brief the types of Naïve Bayes Classifier ?

Q.7 Write the code for adding Naïve Bayes Classifier from sklearn library?