Naive Bayes classifier Algorithm

Based on Bayes theorem

mainly used in text classification includes high dimensional training dataset.

Fast ml model helps in quick predictions.

Probabilistic classifier — predicts based on probability

Applns:

1) Spam filtration 2) sentimental Analysis

3) classifying Articles.

Naive It assumes that the occurrence of a certain feature is independent of the occurrence of other features.

Such a fruit is identified on the basis of color, shape and taste, then red, spherical fruit (apple) without dep on other features.

Bayes — depends on Baye's theorem

Baye's theorem

Known as Baye's Rule or Baye's law.

used to determine the probability of the hypothesis with prior knowledge.

depends on conditional probability

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

$P(A|B)$ - posterior prob - prob of hypothesis A on observed event B

$P(B|A)$ - Likelihood prob - prob of evidence given that the probability of a hypothesis is true.

$P(A)$ - Priori prob - prob of hypothesis before observ evidence

$P(B)$ - Marginal prob - prob of Evidence.

Working on Naive Baye's classifier

We have a dataset of "weather conditions" and corresponding target variable "play". We have to decide wheter we can play on that day or not

1) convert given dataset into frequency table.
2) Generate likelihood table by finding the probabilities of given features.
3) use Bayes theorem, calculate posterior probability.

Problem: If weather is sunny, then player should play or not?

S.NO	outlook	play
0	Rainy	Yes
1	Sunny	Yes
2	overcast	Yes
3	overcast	No
4	Sunny	Yes
5	Rainy	Yes
6	Sunny	Yes
7	overcast	No
8	Rainy	No
9	Sunny	Yes
10	Sunny	No
11	Rainy	Yes
12	overcast	Yes
13	overcast	

Frequency table

overcast	Yes	No
Sunny	3	2
Rainy	2	2
Total.	10	5

Likelihood table Weather conditions

Weather	No	Yes	
Overcast	0	5	5/14 = 0.35
Rainy	2	2	4/14 = 0.29
Sunny	2	3	5/14 = 0.35
All	4/14 = 0.29	10/14 = 0.71	

Applying Bayes theorem

$$P(\text{Yes} \mid \text{sunny}) = \frac{P(\text{sunny} \mid \text{yes}) * P(\text{yes})}{P(\text{sunny})}$$

$$= \frac{3/10 * 10/14}{5/14} = \frac{0.3 * 0.71}{0.35}$$

$$P(\text{yes} \mid \text{sunny}) = 0.60$$

$$P(\text{NO} \mid \text{sunny}) = \frac{P(\text{sunny} \mid \text{NO}) * P(\text{NO})}{P(\text{sunny})}$$

$$= \frac{2/4 * 0.29}{5/14} = \frac{0.5 * 0.29}{0.35}$$

$$P(\text{NO} \mid \text{sunny}) = 0.41$$

$P(\text{yes} \mid \text{sunny}) > P(\text{NO} \mid \text{sunny})$ — so (play)