Day 6

Agenda:

- 1. Naive Bays Algorithm.
- 2. Performance matrix for classification
 - 1. Confusion Matrix
 - 2. Precesion, Recall, F1 score, AUC, ROC
- 3. Sentiment analysis Project
- 4. Imbalancity treatment
- 5. Churn Prediction Project
- 6. Predict if a person will purchase iPhone or not.
- 7. MCQs

Naive Bays Algorithem

Conditional Probability:

Conditional probability is a measure of the probability of an event occurring given that another event has (by assumption, presumption, assertion, or evidence) occurred.

$$P(A \mid B) = \frac{P(A \cap B)}{P(A \cap B)}$$
Probability of event A given B has occured
$$P(B) = \frac{P(B) = 0}{P(B)}$$
Probability of event B

Bays Theorem

Bayes' Theorem is a simple mathematical formula used for calculating conditional probabilities.

Probability of B occurring given evidence A has already
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probability of A occurring given evidence B has already occurred $P(B|A) \cdot P(A)$

Probability of B occurring given evidence B has already occurred

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

- Y: class of the variable
- X: dependent feature vector (of size n)

Naive Bays

- -- Bayes' rule provides us with the formula for the probability of Y given some feature X. In real-world problems, we hardly find any case where there is only one feature.
- -- When the features are independent, we can extend Bayes' rule to what is called Naive Bayes which assumes that the features are independent that means changing the value of one feature doesn't influence the values of other variables and this is why we call this algorithm "NAIVE"

$$P(Y=k|X1,X2...Xn) = \frac{P(X1|Y=k)*P(X2|Y=k)....*P(Xn|Y=k)*P(Y=k)}{P(X1)*P(X2)...*P(Xn)}$$
 This formula can also be understood as
$$\frac{\text{Class Prior Probability}}{P(c|x) = \frac{P(x|c)P(c)}{P(x)}}$$
 Posterior Probability
$$\frac{P(x|x)P(x)P(x)}{P(x)}$$
 Predictor Prior Probability

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

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

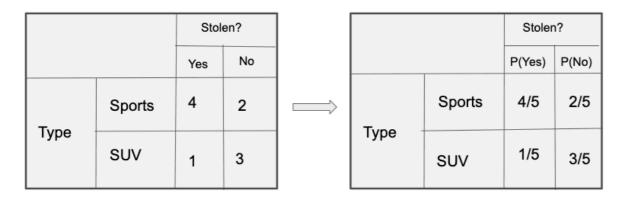
Frequency Table

Likelihood Table

		Stol	en?			Stoler	1?
		Yes	No			P(Yes)	P(No)
	Red	3	2		Red	3/5	2/5
Color	Yellow	2	3	Color	Yellow	2/5	3/5

Frequency Table

Likelihood Table



Frequency Table

Likelihood Table

		Stole	en?			Stoler	1?
		Yes	No			P(Yes)	P(No)
	Domestic	2	3	0.1-1-	Domestic	2/5	3/5
Origin	Imported	3	2	Origin	Imported	3/5	2/5

Color	Туре	Origin	Stolen
Red	SUV	Domestic	?

As per the equations discussed above, we can calculate the posterior probability $P(Yes \mid X)$ as:

P(Yes | X) = P(Red | Yes) * P(SUV | Yes) * P(Domestic | Yes) * P(Yes)

= 3/5 * 1/5 * 2/5 * 1 = 0.048

and, P(NoIX):

P(No | X) = P(Red | No) * P(SUV | No) * P(Domestic | No) * P(No)

= 3/5 * 3/5 * 3/5 * 1 = 0.144

Since 0.144 > 0.048, Which means given the features RED SUV and Domestic, our example gets classified as 'NO' the car is not stolen.



<	3 totaa v		
So in our example, we h	ave 5 predictors X.		