Naive Bayes Classifier

- Its a probabilistic machine learning model that's used for classification task.
- It is based on the Bayes theorem.

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

where:

P(A|B) = Conditional Probability of A given B
P(B|A) = Conditional Probability of A given B
P(A) = Probability of event A
P(B) = Probability of event A

Types of Naive Bayes Classifier:

Multinomial Naive Bayes:

This is mostly used for multiclass.

Bernoulli Naive Bayes:

The parameters that we use to predict the class variable take up only values yes or no, for example if a word occurs in the text or not.

Gaussian Naive Bayes:

When the predictors take up a continuous value and are not discrete.

Eg.

$$X = (x_1, x_2, x_3, ..., x_n)$$

$$P(y|x_1, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

Assignment:

1) change the parameters in logistic reg with titanic dataset (train test split, normalization, select different attribute)

2) Naive bayes (try to apply feature scaling & train test split)

3) Apply Naive Bayes algo on titanic dataset.

4) apply logistic regression on bank dataset.

5) Apply LR,NB on Iris dataset.

Naive Bayes Example

Outlook

	Y	N	P(Y)	P(N)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Tot	9	5	100%	100%

Temp

	Y	N	P(Y)	P(N)
Hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Tot	9	5	100%	100%

Play

Y	9	9/14
N	5	5/14
Tot	14	

Person will Play or not

Today (Sunny, Hot) = Play = ?

x_1 x_2

$$P(\overset{A}{Y}/\overset{B}{Today}) = \frac{P(Sunny/yes)\,P(Hot/yes)\,P(yes)}{P(Today)}$$

$$= \frac{2/9 \times 2/9 \times 9/14}{5/14 \times 4/14} = 0.31$$

$$P(N/Today) = \frac{3/5 \times 2/5 \times 5/14}{5/14 \times 4/14} = 0.84$$

$$P(yes) = \frac{0.31}{0.31 + 0.84} = 0.26$$

$$P(N) = 1 - 0.26 = 0.74$$

Hence
Play = No

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

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

Frequency Table		Play	
		Yes	No
Humidity	High	3	4
	Normal	6	1

Frequency Table		Play	
		Yes	No
Wind	Strong	6	2
	Weak	3	3

Calculating likelihood of each attribute

Likelihood Table		Play		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	3/9	2/5	5/14
		10/14	4/14	

$P(B|A) = P(Sunny|Yes) = 3/9 = 0.33$

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

$P(A) = P(Yes) = 10/14 = 0.71$

Similarly likelihood of "No" given Sunny is:

$P(A|B) = P(No|Sunny) = P(Sunny|No)^* P(No) / P(Sunny) = (0.4 \times 0.36) / 0.36 = 0.40$

Likelihood table for Humidity

Likelihood Table		Play		
		Yes	No	
Humidity	High	3/9	4/5	7/14
	Normal	6/9	1/5	7/14
		9/14	5/14	

Likelihood table for Wind

Likelihood Table		Play		
		Yes	No	
Wind	Weak	6/9	2/5	8/14
	Strong	3/9	3/5	6/14
		9/14	5/14	

$P(Yes|High)$ = 0.33 x 0.6 / 0.5 = 0.42

$P(No|High)$ = 0.8 x 0.36 / 0.5 = 0.58

$P(Yes|Weak)$ = 0.67 x 0.64 / 0.57 = 0.75

$P(No|Weak)$ = 0.4 x 0.36 / 0.57 = 0.25

Outlook = Rain
Humidity = High
Wind = Weak
Play = ?

Likelihood of "**Yes**" = P(Outlook = Rain|**Yes**)*P(Humidity= High|**Yes**)* P(Wind= Weak|**Yes**)*P(**Yes**) = 2/9 * 3/9 * 6/9 * 9/14 = 0.0199

Likelihood of "**No**" = P(Outlook = Rain|**No**)*P(Humidity= High|**No**)* P(Wind= Weak|**No**)*P(**No**) = 2/5 * 4/5 * 2/5 * 5/14 = 0.0166