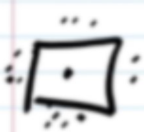
Naive Bayes

→ Supervised Machine Learning Algorithm

→ Based on Conditional Probability.

Probability → What is the Probability that a dice will give me 3 in a roll?

$$P(3) = \frac{1}{6}$$



$$\text{Probability} = \frac{\text{Number of desired Outcomes}}{\text{Total no. of Outcomes}}$$

Rolling a dice = [1, 2, 3, 4, 5, 6]

⇒ Total no. of Outcomes = 6

Conditional Probability

$P(A|B)$ → Probability of A condition B

→ Probability that A occurs given that B has already occured.

$$P(\underbrace{D_1 = 2}_{A} \,|\, \underbrace{D_1 + D_2 \le 5}_{B})$$

Dice 2

Dice 1	1	2	3	4	5	6
1	×	×	×	×		
✓ 2	× •	× •	× •			
3	×	×				
4	×					
5						
6						

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

$$P(B) = \frac{10}{36} \quad ; \quad P(A \cap B) = \frac{3}{36}$$

$$P(D_1 = 2 \mid D_1 + D_2 \leq 5) = \frac{3}{10}$$

$$\boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)}} \rightarrow \text{Conditional Probability}$$

(1) $P(A \mid B) \times P(B) = P(A \cap B)$

(2) $P(B \mid A) \times P(A) = P(A \cap B)$

$P(A \mid B) \times P(B) = P(B \mid A) \times P(A)$

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

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

Bayes Theorem

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

After Derivative

$$\left[P(C_k \mid x) \propto P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k) \right] \rightarrow \text{Naive Bayes}$$

Naive Bayes algorithm is basically used for text classificatn problems.

e.g. Spam or Ham

Text	Target
→ You, won 1 million.	Spam
→ Hello Ram, Come for meeting.	ham
→ Donation of ₹10,000 to your account	Spam

Test data.

↑

x → The Queen wants to send you $10,000. →

$$\underbrace{P(C_{spam} \mid x)} > P(C_{ham} \mid x)$$

$\rightarrow$ This is a spam mail.

$$\begin{cases} \Rightarrow P(C_{spam} \mid x) \propto P(C_{spam}) \times P(\text{The} \mid C_{spam}) \times P(\text{Queen} \mid C_{spam}) \\ \qquad \times P(\text{wants} \mid C_{spam}) \times P(\text{to} \mid C_{spam}) \times \ldots \times \\ \qquad P(\$10,000 \mid C_{spam}) \\ \Rightarrow P(C_{ham} \mid x) \propto P(C_{ham}) \times P(\text{The} \mid C_{ham}) \times P(\text{Queen} \mid C_{ham}) \\ \qquad \times \ldots \times P(\$10,000 \mid C_{ham}) \end{cases}$$

compared

$$\left. \begin{array}{l} P(\text{The} \mid C_{spam}) = 0.32 \\ P(\text{Queen} \mid C_{spam}) = 0.12 \end{array} \right\} \rightarrow \text{Calculated and Stored in Training Phase.}$$

$\vdots$

Test

$x \rightarrow$ You won ₹10,000

$P(\text{You} \mid C_{ham})$ $\quad$ $P(\text{won} \mid C_{ham})$ $\quad$ $P(\text{₹}10{,}000 \mid C_{ham})$

$P(\text{You} \mid C_{spam})$ $\quad$ $P(\text{won} \mid C_{spam})$ $\quad$ $P(\text{₹}10{,}000 \mid C_{spam})$

$$P(C_{ham} \mid x) = P(C_{ham}) \times \_\_\_ \times \_\_\_ \times \_\_\_ = 0.4$$

$$P(C_{spam} \mid x) = P(C_{spam}) \times \_\_\_ \times \_\_\_ \times \_\_\_ = 0.6$$

x belongs to spam category.

Bag of Words → Every word will become a feature.

→ I am a good boy.
→ I am a bad boy.
→ ⋮
→ ⋮
→ 10,000 data point.

I	am	a	good	boy	bad
1	1	1	1	1	0
1	1	1	0	1	1

e.g →

	Dear	Friend	Lunch	Money	(Label) Ham/Spam
1	0	1	0	2	Spam
2	1	1	0	2	Spam
3	0	1	2	0	Spam
4	1	1	1	1	Ham -
5	0	1	1	1	Ham -

1) Friend money money
2) Dear Friend Money money
3) Friend Lunch Lunch
4) Dear Friend Lunch Money

5) Friend lunch money.

$P(\text{spam}) = \frac{3}{5}$; $P(\text{ham}) = \frac{2}{5}$

Training Phase

$P(\text{Dear} \mid \text{Spam}) = \frac{1}{3} = 0.3$ $\quad P(\text{Dear} \mid \text{ham}) = \frac{1}{2} = 0.5$

$P(\text{Friend} \mid \text{spam}) = \frac{3}{3} = 1$ $\quad P(\text{Friend} \mid \text{ham}) = \frac{2}{2} = 1$

$P(\text{Lunch} \mid \text{Spam}) = \frac{2}{3} = 0.6$ $\quad P(\text{lunch} \mid \text{ham}) = \frac{2}{2} = 1$

$P(\text{money} \mid \text{Spam}) = \frac{4}{3} = 1.3$ $\quad P(\text{money} \mid \text{ham}) = \frac{2}{2} = 1$

New Test Point

$x \rightarrow$ lunch money money

$$P(\text{spam}/x) \rightarrow P(\text{spam}) \times P(\text{lunch} \mid \text{spam}) \times P(\text{money} \mid \text{spam}) \times P(\overset{\text{money}}{\text{mny}} \mid \text{spam})$$

$$P(\text{ham}/x) \rightarrow P(\text{ham}) \times P(\text{lunch} \mid \text{ham}) \times P(\text{money} \mid \text{ham}) \times P(\text{money} \mid \text{ham})$$

$P(spam|x) \rightarrow 0.6 \times 0.6 \times 1.3 \times 1.3 = 0.6084$ | 0.4

$P(ham|x) \rightarrow 0.4 \times 1 \times 1 \times 1 = 0.4$ | $\boxed{0.6}$

$x \rightarrow$ lunch money money belongs to <u>spam</u> category.

Q) What if a word comes which is not present in the training phase?

Let say $x \rightarrow$ You have won 1 lakh, click here to claim your prize.

$$P(spam|x) \rightarrow P(spam) \times \boxed{\cancel{P(You|spam)}}^{\rightarrow 0} \times P(have|spam) \times \ldots\ldots$$
$$\times P(prize|spam) = 0$$

Similarly,

$$P(ham|x) \rightarrow P(ham) \times \cancel{P(You|ham)}^{\rightarrow 0} \times \ldots \times P(prize|ham)$$
$$= 0$$

$$P(\text{You} \mid \text{spam}) = \frac{\text{Prob. of word You occurs and it's spam}}{\text{Probability of Spam}}$$

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

$$= \frac{0}{\text{Prob. of Spam}} = 0$$

In order to control this type of Situation, we add a term called α in Numerator and $K\alpha$ in denominator.

$$= \frac{0 + \alpha}{\text{Prob of Spam} + K\alpha} \neq 0 \; ; \quad \alpha \neq 0$$

This process of adding α and $K\alpha$ is known as Laplacian Smoothing.

⊗ Why naive Bayes is so naive?

→ It assumes that all the features are independent in nature.

When to use fit_transform and transform?

→ fit → learn the data

→ transform → Apply the data which was learned by fit.

→ fit_transform → learn the data and applies the data.

fit_transform → Training Data

transform → Test Data

If you apply fit_transform on all of your 'X' there will be some data leakage.

e.g.

Scaling

Min-Max Scaler()

X	
22	–
44	–
66	–
88	–
11	–

→ fit_transform (x)

→ mixed data.

→ [0.2, 0.4, 0.6, 1, 0.1]

→ split

→ Always split your data and apply the following ↑.

TF-IDF → Term Frequency - Inverse Document Frequency.

Imp → More weightage to lesser occuring words and lesser weightage to the most occuring word.

$$\text{Term Frequency} = \frac{\text{Number of times word appeared in document}}{\text{Total no. of words in document}}$$

$$\text{Idf} = \log\left(\frac{\text{Total number of document}}{\text{No of document where the word has come.}}\right)$$

→ 100 words

[cat, dog, cat, cat, dog]

[ ]

[ ]

Documents

→ Corpus

e-g Document → 100 words → word "cat" appeared 3 times

$$\text{tf (term Frequency)} = \frac{3}{100} = 0.03$$

e.g. 10 million documents → word "cat" appeared in 1000 documents,

$$\log\left(\frac{10,000,\cancel{000}}{1\cancel{000}}\right) = \log(10,000) = 4$$

$$tf \times idf = 4 \times 0.03 = 0.12$$

[is is is]
[is]
[is]
[is]
⋮
10 million documents

$$0.03 \times \log\left(\frac{10,000,000}{10,000,000}\right)$$

$$0.03 \times \overset{0}{\log}(1)$$

$$\Rightarrow \underline{\underline{0.03 \times 0.}} = 0$$