Points about Naive Bayes :-

→ Naive Bayes learn parameters by looking at each feature individually and collect simple per-class statistics from each feature.

→ Three kinds of Naive Bayes classifier implemented in scikit-learn i.e GaussianNB, BernouliNB, MultinomialNB.

GaussianNB :-

→ It can be applied to any continuous data.

BernouliNB :-

→ It assumes binary data. ~~The~~ It count how often every feature of each class is not-zero.

MultinomialNB :-

→ It assumes count data (i.e, that each features represents an integer count of some-thing, like how often a word appears in a sentence.)

⊛ BernouliNB & MultinomialNB are mostly used in text data classification.

⊛ MultinomialNB takes into account the average value of each feature for each class.

⊛ GaussianNB stores the average value as well as standard deviation of each feature for each class.

Strengths, Weakness & Parameters :-

→ MultinomialNB & BernouliNB have a single parameter, alpha, which control Model Complexity.

Working of Alpha :-

→ The way alpha works is that the algorithm adds to the data alpha many virtual data points that have positive values for all features.

→ This results in a "smoothing" of the statistics. A large alpha means more smoothing, resulting in less complex models.

⊛ The algorithm's performance is relatively robust to the setting of alpha, meaning that setting alpha is not critical for good performance. However tuning it usually improves accuracy somewhat.

→ GaussianNB is mostly used on very high dimensional data.

→ MultinomialNB & BernoulliNB → used for sparse data i.e (text)

→ Multinomial NB is better than BernoulliNB. (on sparse data)

$\boxed{\rightarrow}$

→ Very fast to train & to predict.

→ Training procedure is easy to understand.

→ Models work very well with high-dimensional sparse data.

→ Relatively Robust to the parameters.

⊛ Naive Bayes models are great baseline models. and are often used on very large datatsets, where training even a linear model might take too long.

- **Naive Bayes :-**

1. CONDITIONAL PROBABILITY :-

$A, B \to$ Events

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

$\cap \to$ Intersection

$\downarrow$ Given

given, $\boxed{P(B) \neq 0}$

eg:- Two Dices rolling together.

D_1, D_2

$\{(1,1), (1,2) \ldots (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\} \to$ 36 Sample space

D1

D2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

i) - What is the probability the value come 5 in D_1.

$\cancel{P(A)}$. $P(A=5) = \frac{1}{36}$

ii) - What is the probability $D_1 + D_2 \leq 10$

$P(B) = P(D_1 + D_2 \leq 10) = \frac{33}{36} = \frac{11}{12}$

iii) - Probability of $D_1 = 5$ given $D_1 + D_2 \leq 10$

$P(D_1 = 5 | D_1 + D_2 \leq 10)$

$P(A|B) = \frac{5}{33}$

$A = \{D_1 = 5\}$, $B = \{D_1 + D_2 \le 10\}$

$P(A \cap B) = \frac{5}{36}$

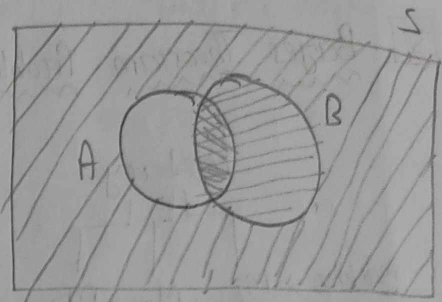
$P(B) = \frac{33}{36}$

$\therefore \; P(A|B) = \frac{5/36}{33/36} = \frac{5}{33}$

2 Independent Events :-

$A, B \rightarrow$ Events

$$\boxed{P(A \cap B) = P(A) * P(B)}$$



$$\boxed{P(A|B) = \frac{n(A \cap B)}{n(B)}} \quad -(i)$$

$$\Rightarrow P(A) = \frac{P(A \cap B)}{P(B)} = \frac{n(A \cap B)}{n(S)} * \frac{n(B)}{\cancel{n(A \cap B)}}$$

$n(S)$

$$\boxed{P(A) = \frac{n(A \cap B)}{n(B)}} \quad -(ii)$$

from (i) and (ii)

$$\Rightarrow \boxed{P(A|B) = P(A)}$$

3 Mutually Exclusive Events :-

$A, B \rightarrow$ Events

$$\boxed{P(A \cap B) = 0}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow \boxed{P(A|B) = 0}$$

4 Bayes Theoram :-

A, B → Events

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

likehood → $P(B|A)$; Prior → $P(A)$; Evident → $P(B)$

Posterior → $P(A|B)$ Given $\boxed{P(B) \neq 0}$

5 Bayes Theoram Problem :-

Factory

	M_1	M_2	M_3
machine →	M_1	M_2	M_3
made →	(20%)	(30%)	(50%)
Defect manufac →	(5%)	(3%)	(1%)

Q.

- Batch of marker. Randomly choose a marker and that is defective. What is the probability that this marker comes from M_3.

Ans:-

$P(M_1) = \frac{20}{100} = \frac{1}{5}$ $P(D|M_1) = \frac{5}{100} = \frac{1}{20}$

$P(M_2) = \frac{30}{100} = \frac{3}{10}$ $P(D|M_2) = \frac{3}{100}$

$P(M_3) = \frac{50}{100} = \frac{1}{2}$ $P(D|M_3) = \frac{1}{100}$

Marker ↓
Defective → Yes
Probability → M_3

- B i.e $P(M_3|D) = \frac{P(D|M_3) * P(M_3)}{P(D)} = \frac{\frac{1}{100} * \frac{1}{2}}{P(D)}$

$\Rightarrow P(D) = P(D \cap M_1) + P(D \cap M_2) + P(D \cap M_3)$ we know $P(A|B) = \frac{P(A \cap B)}{P(B)}$

$\Rightarrow P(D) = P(D|M_1) * P(M_1) + P(D|M_2) P(M_2) + P(D|M_3) * P(M_3)$

$\Rightarrow P(D) = \frac{1}{20} \times \frac{1}{05} + \frac{3}{100} \times \frac{3}{100} + \frac{1}{100} \times \frac{1}{2} = \frac{1}{100} + \frac{9}{1000} + \frac{1}{200} =$

7 Mathematics behind Naive Bayes Algorithm :-

Let us assume,

$X = \{x_1, x_2, x_3 \dots x_n\}$

$C_K = \{C_1, C_2, C_3 \dots C_K\}$ → Target

multiclass classification

x_1 toss	x_2 Venue	x_3 outlook	result (Binary) Won / loss C_1 C_2

We know that

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

↳ This term is common, so remove this.

$\Rightarrow P(C_K|X) = P(X|C_K) P(C_K)$

$\Rightarrow P(C_K|X) = P(X \cap C_K)$ ← $\boxed{\cap \to ,}$

We know, conditional Probability

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

$\Rightarrow P(A|B) P(B) = P(A \cap B)$

$\Rightarrow P(C_K|X) = P(X, C_K)$

$A \cap B = P(A|B) P(B)$ ⇒ [Chain rule for conditional Probability]

$= P(\boxed{x_1}, \boxed{x_2, x_3, \dots x_n, C_K})$

$= \underbrace{P(x_1 | x_2, x_3, \dots x_n, C_K)}_{a} P(x_2, x_3 \dots x_n, C_K)$

$A \cap B = P(A|B) P(B)$

$\Rightarrow P(C_K|X) = a \times P(\boxed{x_2} \boxed{x_3, \dots, x_n, C_K})$

$= a \times \underbrace{P(x_2 | x_3, x_4 \dots x_n, C_K)}_{b} P(x_3, x_4 \dots x_n, C_K)$

$= ab\, P(x_3, x_4, \dots x_n, C_K)$ a

$\boxed{\text{* } P(x_n | C_K) P(C_K) \text{ (in last term)}}$

$= P(x_1 | x_2, x_3 \dots x_n, C_K) P(x_2 | x_3, x_4 \dots x_n, C_K) \dots P(x_{n-1} | x_n, C_K)$ $P(x_n | C_K) P(C_K)$

* Conditional Independence i.e $P(A|B, C) = P(A|C)$ (apply Naive assumption)

$= P(x_1|C_K) P(x_2|C_K) P(x_3|C_K) \dots P(x_{n-1}|C_K) P(C_K)$

8

$$\boxed{P(C_k \mid x) = \frac{1}{Z} P(C_k) \prod_{i=1}^{n} P(x_i \mid C_k)}$$

Product

Z = Probability of x.

$$\boxed{\hat{y} = \underset{k \in \{1,2,\ldots,K\}}{\arg\max} \; P(C_k) \prod_{i=1}^{n} P(x_i \mid C_k)}$$

Maximum a postiorior Rule

MAP

8 Handling Numerical Data :-

Height	Weight	Gender
172	150	M
180	170	M
165	140	M
190	200	M
139	100	F
145	120	F
166	140	F
172	150	F

* Gaussian Naive Baye — sklearn

Q.

H = 185, W = 170, G = ?

Total datapoint = 8

* Take assumtion

$P(M \mid H=185, W=170) = P(H=185 \mid M) \times P(W=170 \mid m)\; P(m)$

$\frac{1}{2}$

=> the height is a Gaussian distributed random variable.

i). Mean. and std of male datapoint.

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}}$$

x = 185 for H, m dp = 4

x = 170 for W F dp = 4

find out probability to particular happening

(*) Not normal distributed then use → Binomial, Multinomial, Possion Naive Bayes

6 ~~Naive Bayes~~ Naive Bayes — General Intuition :—

CSK

toss	Venue	Outlook	Result
won	Mumbai	overcast	won
lost	Chennai	sunny	won
won	Kolkata	sunny	won
won	Chennai	sunny	won
lost	Mumbai	sunny	lost
won	Chennai	overcast	Lost
won	Kolkata	overcast	lost
won	mumbai	sunny	won

input — output

→ Binary classification problem

Total = 8 datpoints

Q. x_q = { lost, mumbai, sunny } → Condition

Result = ? {W | L} {lost ∩ mumbai ∩ sunny}

→ P(W | Condition) = P($\overset{A}{W}$ | $\overbrace{\text{lost} \cap \text{mumbai} \cap \text{sunny}}^{B}$)

P(L | condition) = P(L | lost ∩ mumbai ∩ sunny) = ~~3/10~~

We know that, Bayes Theorem, $P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$

~~P(W|~~ = $\frac{P(\text{lost} \cap \text{mumbai} \cap \text{sunny} \mid P(W)) * P(W)}{P(\text{lost} \cap \text{mumbai} \cap \text{sunny})}$

(Same cases)

= (P(lost ∩ mumbai ∩ sunny) | P(W)) * P(W), P(W) = 5/8

0 — 5/8 — P(L) = 3/8

= 0

But, ⇒ P(lost |W) P(Mumbai |W) P(sunny |W) * P(W)

= 1/5 × 2/5 × 4/5 × 5/8 = 80/120 = 2/3