Naive Bayes :- Before going to naive bayes algorithm, here we

have to understand Some topics of Statistics.

① Random Experiment :- it is a process, where Outcomes doesn't predict with Certainity (ie Contain Uncertinity)

Ex : Tossing a coin, Rolling Dice

② Sample Space :- Output of Random Experiment is called as Sample Space.

Ex : Rolling dice – {1, 2, 3, 4, 5, 6}, Tossing Coin – {H, T}.

③ Event :- Subset of Sample Space.

Here we are finding out the probabilities of events, for Suppose consider R.D then what is the probability of getting on odd Number?

R.E → Rolling Dice

S.S → {1, 2, 3, 4, 5, 6}

Event → {1, 3, 5}

$$P.(\text{Event}) \rightarrow \frac{\text{Outcomes favarable}}{\text{Total No of possibility}} = \frac{3}{6} = \frac{1}{2} \Rightarrow 0.5 \Rightarrow 50\% \text{ probability}$$

∴ from probability of event we can say that there is 50% of chances to get odd number in Rolling a dice

from the Same event given that prime number appeared on the dice rolling ? what is the probability of it ?

So in this case we have two Events appeared. Event$_1$ ⇒ E_1

E_1 = Getting odd Number ; E_2 = Appear prime Number.

R.E → Rolling a dice

S.S → $\{1, 2, 3, 4, 5, 6\}$

$E_1 \rightarrow \{1, 3, 5\}$; $E_2 \rightarrow \{2, 3, 5\}$

from this we can — $\boxed{P(E_1/E_2) = \dfrac{P(E_1 \cap E_2)}{P(E_2)}}$ Conditional probability.

$$P(E_1 | E_2) = \frac{2/6}{3/6} = \frac{2}{3}$$

both are Subsets

Independent Events :- $P(E_1/E_2)$ ⇒ Says probability of E_1 getting Given that E_2 already occured, if E_1 is independent of E_2, then the probability of event (E_1) not going to Impact presence of E_2.

From this we can Say No Impact of E_2 on E_1 then Simply —

$$\boxed{\begin{aligned} P(E_1|E_2) &= P(E_1) \\ P(E_2|E_1) &= P(E_2) \end{aligned}}$$ Similarly

Conditional probability :- for this above eqns we can Say that Conditional probability of events — $\boxed{P(E_1|E_2) = \dfrac{P(E_1 \cap E_2)}{P(E_2)}}$ — ① ∵ $\boxed{E_1 \cap E_2 = E_2 \cap E_1}$

Bothare Same.

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \quad \text{— ②}$$

Eq (2) can written as — $\boxed{P(E_1 \cap E_2) = P(E_2|E_1) \cdot P(E_1)}$ — ③

if we Substitute Eq (3) in Eq (1) Then —

$$\text{posterior} \leftarrow \boxed{P(E_1|E_2) = \frac{P(E_2|E_1) \cdot P(E_1)}{P(E_2)}} \quad \text{— ④}$$

Likelyhood

Prior

Evidence , $P(E_2 \neq 0)$

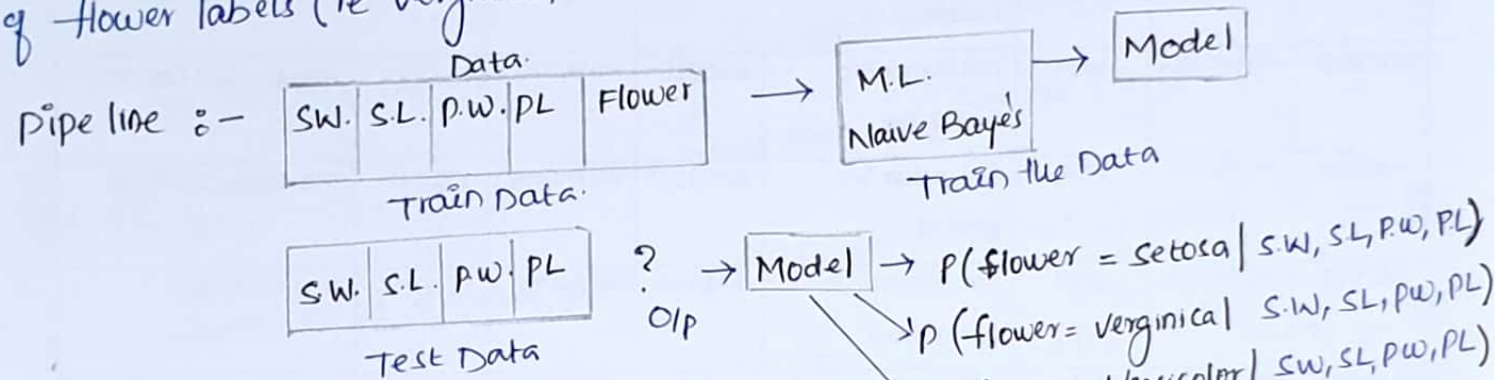
Eq (4) is called as Bayes Theorem.

Naive Baye's Algorithm :-

Based On Bayes Theorem, we derive Naive Baye's Algorithm. it is a Supervised Technique in Machine learning, from this we solve problems only classification (ie O/P → Categorical / Descrite type)

Let Consider a dataset of iris in that we have input variables as (Splwidth, Spl length, petal width, petal length), ~~form~~ Output variable as flower from the given input data, we predict the Label of flower (ie Output) of x_q (future query point) by Using Naive bay's Algorithm. in which we Caliculate the probability of flower labels (ie Verginica, Setosa, Versicolor)

pipe line :-



from the probabilities of flowers — $P(S) = 0.7$ — highest probability

$P(Ver) = 0.2$

$P(VC) = 0.1$

We caliculate probabilities of each flower in presence of given input we caliculate the Label of flower from this Setosa Contain highest probability so the future query point (x_q) belonging to "Setosa"

Task :- Given x_q, find the probability of x_q belonging to class C_i

where $i \in (1, 2, 3 \ldots k) \Rightarrow P(C_k / x_q)$.

Mathematic Intiution in Naive Baye's :-

from Naive Baye's We are intrestead to find Out the Probability of $P(C_k|x_q)$, Accoding to Baye's therom we Can represent

as –
$$\boxed{P(C_k|x_q) = \frac{P(x_q|C_k) \cdot P(C_k)}{P(x_q)}} \quad \text{— eq ①}$$

where – $x_q \rightarrow$ Given future or Query point.

$C_k \rightarrow$ Given Spreces (setosa, vercica, vergenica)

$P(C_k|x_q) \rightarrow$ from given Query point (x_q) we are finding out probability (C_k) belonging to which Label output.

So in iris data set we have ③ output Label value So we consider them as (C_1, C_2, C_3) if there are more labels then we Consider $(C_1, C_2 - - - C_k)$

from eq (1) we can write – $P(C_1|x_q) = \frac{P(x_q|C_1) \cdot P(C_1)}{x_q}$ — ②

$$P(C_2|x_q) = P(x_q|C_2) \cdot P(C_2) / x_q \quad — ③$$
$$P(C_3|x_q) = P(x_q|C_3) \cdot P(C_3) / x_q \quad — ④$$

if we observe eqs ②,③,④ the denominator (x_q) is Same in all classes. So if we remove the (x_q) then eq will be represent as –

$$\boxed{P(C_k/x_q) \propto P(x_q/C_k) \cdot P(C_k)} \quad — ⑤$$

Probability of (C_k) given on x_q is directly prapotional to probability of (x_q) given on C_k * probability of C_k ie the given (x_q) future point is not belonging to any of Label class. as per eqn ⑤.

from Conditional probability we can write –

$P(x_q|C_K) \cdot P(C_K) = P(x_q \cap C_K)$ so put this in eq 5 Then –

$$P(C_K|x_q) \propto P(x_q \cap C_K)$$

Now $P(x_q \cap C_K)$ can be written as – $P(x_q, C_K)$ Now eqn will be –

$$\boxed{P(C_K/x_q) \propto P(x_q, C_K)} \quad — ⑥$$

According to chain rule of Conditional probability, the right Side part of eqn ⑥ will be written as –

$$\boxed{P(x_q, C_K) = P(x_1, x_2, x_3, \ldots x_d, C_K)} \quad — ⑦$$

In this we consider x_1 as one part and remaining all as another part and (,) represent intersection between them. $\Rightarrow P(A \cap B) = P(A|B) \cdot P(B)$

Similarly eqn 7 will written as –

$$\boxed{P(\underbrace{x_1}_{A} \cap \underbrace{(x_2, x_3, \ldots x_d, C_k)}_{B}) = \underbrace{P(x_1|x_2, x_3 \ldots x_d, C_K)}_{P(A/B)} \cdot \underbrace{P(x_2, x_3, \ldots x_d, C_K)}_{P(B)}} \quad - ⑧$$

According to chain rule of Conditional probability, In eq ⑧ the part of $P(B)$ will be expandable till end i.e. $P(x_d)$.

$$\boxed{\begin{aligned} P(x_q, C_K) = {} & P(x_1|x_2, x_3, \ldots x_d, C_K) \cdot P(x_2|x_3, x_4, \ldots x_d, C_K) \\ & P(x_3|x_4, \ldots x_d, C_K) \ldots P(x_d|C_K) \cdot P(C_K) \end{aligned}} \quad - ⑨$$

This Algorithm Contain a naive assumption to Simplyfy this equation

The Naive assumption is All the input variables are Indipendent to each other.

from This Naive assumption we can write as –

$$P(x_1/x_2) = P(x_1) \Rightarrow \text{Independent to each other}$$

So if we apply this on eqn (9) Then –

$$\boxed{P(x_q, C_K) = P(x_1|C_K).\,P(x_2|C_K).\,P(x_3|C_K) \ldots P(x_d|C_K).\,P(C_K)} \quad \text{— (10)}$$

finally we conculded that expand of $P(x_q, C_K)$, Now put this value into eqn (6) Then – $P(C_K|x_q) \propto P(x_1|C_K).P(x_2|C_K) \ldots P(x_d|C_K).P(C_K)$

from baye's Theorem we can write as –

$$P(C_K|x_q) = \frac{P(x_1|C_K).\,P(x_2|C_K) \ldots P(x_d|C_K).P(C_K)}{P(x_q)}$$

$\therefore \quad a_1 * a_2 * a_3 * \ldots * a_n = \prod_{i=1}^{n} a_i$ from This we can write as –

$$\boxed{P(C_K/x_q) = \frac{P(C_K).\prod_{i=1}^{d} P(x_i/C_K)}{P(x_q)}}$$

Here $C_K \rightarrow$ No of Label outputs ie (classes) So here x_q the future point depends on maximum of probability of output label $P(C_K)$ So The predicted output ($\hat{y}$) will be maximum. then we can write as – $\hat{y} = \underset{K}{\arg}\max P(C_K|x_q)$

$$\boxed{\hat{y} = \underset{K}{\arg}\,\text{Max}\left\{\frac{P(C_K).\prod_{i=1}^{d} P(x_i/C_K)}{P(x_q)}\right\}} \Rightarrow \text{Naive Baye's classifier}$$

$K \rightarrow$ class

In generally this Naive Baye's eqn Used in classification tasks only, In real world we Used alot in text data, This cla, ssifier is also called as "Maximum a posteriori" Algorithm

from this we can Say which class or label has maximum probabi-lity we Consider the output is that label. for Suppose in Iris data verginica Contain maximum probability as Compare to Other class then we Consider "Verginica" is output of that given query point.

Hyper parameter of Naive baye's is "α" (alpha)

if $\alpha = 0$ $\Rightarrow$ α value is very low $\Rightarrow$ $\alpha\downarrow$ $\Rightarrow$ Naive baye's going to be "Overfit issue"

if α value is very high $\Rightarrow$ $\alpha\uparrow$ $\Rightarrow$ Naive baye's going to "Under-fit" issue

α is also called as _ Laplace Smoothing (or) Additive Smoothing

Generally we have three different types of Naive bayes eqn are —

1. Gaussion Naive Bayes ← if we have Real value data (Numerical)
2. Multinomial Naive Baye's ← if we have text data. we Use (categorical)
3. Bernoli Naive Baye's ← if we have Binary data (0,1) representation