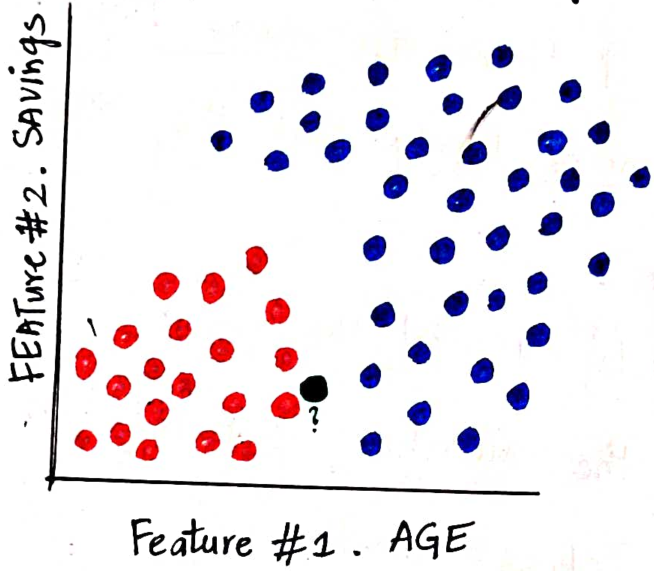
Dt:- 21/05/22

Naive bayes

* Naive Bayes is a classification Technique based on "Bayes Theorm".

Ex:- Lets, assume, You want to classify a new client as eligible to retire (or) not.

* Customer features are his age and salary.



class 1 (Retire) 40

class 0 (not retire) 2

* 1. Prior probability

* Points can be classified as RED or blue

* Our Task is to classify a new data point to RED (or) blue

* Prior probability :- Since, we have more "Blue" compared to red, we can assume that our new point is twice as likely to be Blue Than red.

* Prior Probability for "RED" = $\dfrac{\text{Number of Red points}}{\text{Total no. of points}} = \boxed{\dfrac{20}{60}}$

* Prior Probabity for "blue" = $\dfrac{\text{Number of blue points}}{\text{Total no. of points}} = \boxed{\dfrac{40}{60}}$

2. Likelihood :-

* for the new point, if there are more "Blue" points in its Vicinity, it is more likely that the new point will be classifed as "Blue".

* So, we draw a circle around the point

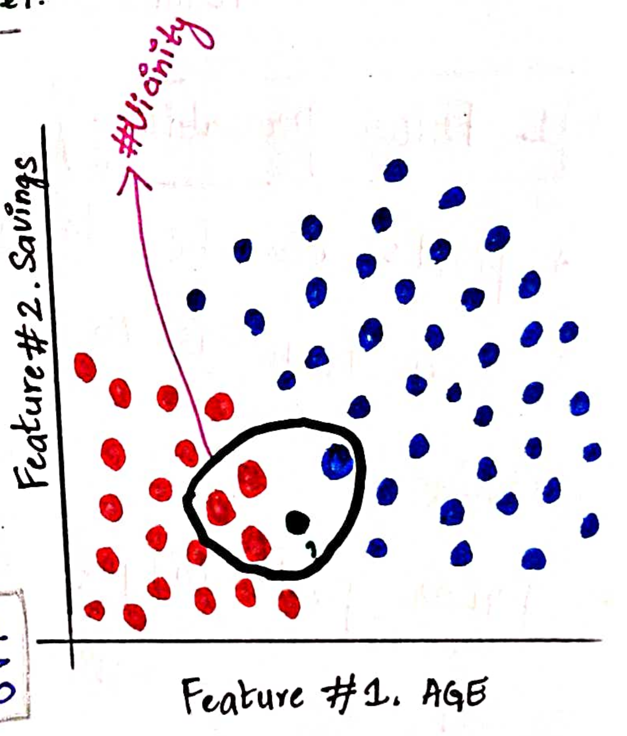
* Then, we calculate the number of points in the circle belonging to each class label.

* Likelihood of being "RED":-

$$= \frac{\text{No. of Red Points in vicinity}}{\text{Total no. of Red Points}} = \boxed{\frac{3}{20}}$$

* Likelihood of being "Blue":-

$$= \frac{\text{No. of blue points in Vicinity}}{\text{total no. of blue Points}} = \boxed{\frac{1}{40}}$$



3. Posterior Probability.

* Let's Combine Prior probability and Liklihood to create a posterior probability.

* Prior probabilities :- Suggests that "X" may be classified as "Blue". Because there are twice as much blue Points.

* Likelihood probabilities :- suggests that "x" is "RED" because there more red points in the vicinity of "x".

* Baye's Rule Combines both to form a Posterior probability.

• Posterior Probability of "x" being "Red" = Prior probability of "RED" * Likelihood of "x" being "RED"

$$= \boxed{\frac{20}{60} \times \frac{3}{20} = \frac{1}{20}}$$

• Posterior probability of "x" being "Blue" =

= Prior probability of "blue" * Likelihood of "x" being "blue"

$$= \boxed{\frac{40}{60} \times \frac{1}{40} = \frac{1}{60}}$$

* **Naive Bayes** :- Naive Bayes is a classification Technique based on "Baye's theorm.

$$\boxed{P(retire|x) = \frac{P(x|retire) * P(retire)}{P(x)}}$$

Conditional Probability

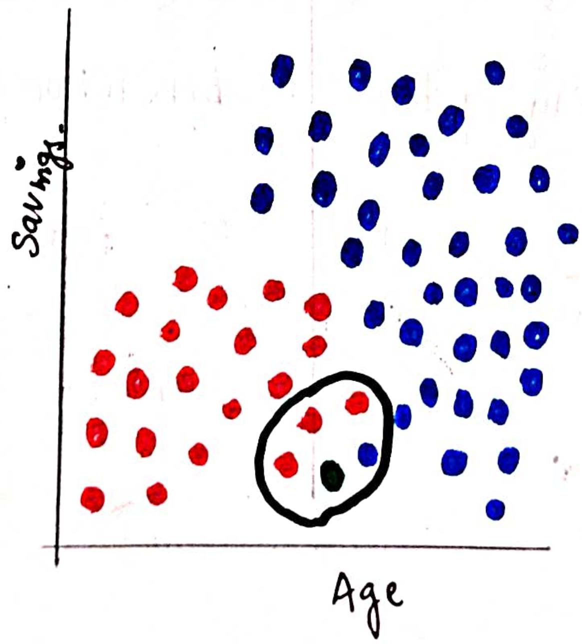
likelihood

Prior probability

$$P(retire|x) = \frac{P(x|retire) * P(retire)}{P(x)}$$

Marginal likelihood

* X → New customer's feature : age & Savings.
* p (retire|x) :- Probability of Customer retire given his Features, Such as age & savings.
* p (retire) :- Prior Probability of retiring, without any prior knowledge.
* P (x|retire) :- Likelihood
* P (x) :- Marginal likelihood, the probability of any point added lies into circle



● → This data points, we want to classify that it red (or) blue data point.

1. Probability of ● data point being ● (blue)

$$P\left[\frac{B}{X}\right] = \boxed{\frac{\text{Prior probability} \times \text{likelihood}}{\text{Marginal likelihood.}}}$$ → fomula

initial data

Actual data point in vicinity

How many data points we considering for analysis

Ex: = 5

$$* \; P\left[\frac{B}{X}\right] = \frac{\frac{\cancel{40}\,(\text{Blue})}{60\,(\text{total})} \times \frac{1\,(\text{Blue})}{\cancel{40}\,(\text{out of total})}}{\frac{4\,(\text{Data points in vicinity})}{60\,(\text{total})}} = \frac{\frac{1}{60}}{\frac{4}{60}} = \frac{1}{4} = \boxed{0.25}$$

For Blue :- ↑

↓

(retire)

*

For Red :- $P\left[\frac{R}{x}\right] = \frac{\frac{\cancel{20}\,(\text{Red})}{\cancel{60}_{3}\,(\text{total})} \times \frac{\cancel{3}\,(\text{red})}{20\,(\text{out of total})}}{\frac{4\,(\text{Data points in vicinity})}{60}} \Rightarrow \frac{\frac{1}{20}}{\frac{4}{60}} = \frac{1}{\cancel{20}_{1}} \times \frac{\cancel{60}^{3}}{4}$

↓

(not retire)

$$= \frac{1}{\cancel{20}_{1}} \times \frac{\cancel{60}^{3}}{4} = \frac{3}{4} = \boxed{0.75}$$

* Sum of probabilities is Equal to = 0.75 + 0.25 = $\boxed{1}$

Q What is Naive ?

A:- it is called "Naive" because it assumes that the presence of certain features in a class is $\boxed{\text{independent}}$ of presence of other features.

* Example : 1 : Age / savings, The assumption is not necessari true. Since age / savings might be dependant on each others.

* Example : 2 : fruit can be classified as watermelon, if its color is green, tastes sweet, and round.

* These Features might be $\boxed{\text{dependent}}$ on each others, however, we assume they are $\boxed{\text{all independent}}$ and that why its "Naive"

We are assuming $\boxed{\text{Every independent Feature}}$ is independent to each other, there is $\boxed{\text{no dependency}}$ between the independent features. This is assumption and we are assuming that features are independent of each other that's the reason it is called as "Naive Bayes"

Que :- what are independent & dependent features ?

A :- Independent features :-

 ⇒ Rolling a "cube" (or) "Dice"

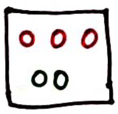
Dice :

$\{1, 2, 3, 4, 5, 6\}$

* Probability :- $1 = \frac{1}{6}$
* Probability :- $3 = \frac{1}{6}$
* Probabity :- $6 = \frac{1}{6}$

independent Events

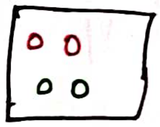
⇒ Dependent features



dependent Variable ←

• First Event

* Probability of picking "red" marble

$P(R) = \frac{3}{5}$ ∘ (remaing) 

• Second Event

* probabilits of picking "Green" marble

$P(G) = \frac{\cancel{2}}{\cancel{4}} = \frac{1}{2}$

$$\boxed{P(\text{Red and Green}) = \text{Probability}(\text{Red}) * \text{Probility}\left(\frac{\text{Green}}{\text{red}}\right)}$$

↳ Conditional probability.

$$P(A \text{ and } B) = P(A) * P(B/A)$$

We write,

$$P(A \text{ and } B) = P(B \text{ and } A)$$

$$P(A) * P\left(\frac{B}{A}\right) = P(B) * P(A/B)$$

$$\boxed{P(B/A) = \frac{P(B) * P(A/B)}{P(A)}}$$ → Baye's Theorm.

$\underbrace{\qquad\qquad\qquad\qquad}$ independent (A)

X_1	X_2	X_3	X_4	$X5$	. . . X_n	y
–	–	–	–	–	–	–

→ output (B)

$$P(y/x_1, x_2, x_3, x_4 \ldots x_n) = \frac{P(y) * P(x_1, x_2, x_3 \ldots x_n)/y}{P(x_1 x_2 x_3 x_4 \ldots x_n)}$$

We expand equation ...

$$= \frac{P(y) * P(x_1/y_1) * P(x_2/y_2) * P(x_3/y_3) \ldots P(x_n/y_n)}{P(x_1) + P(x_2) + P(x_3) + \ldots P(x_n)}$$

Example :-

Dataset

x_1	x_2	x_3	x_4	y
–	–	–	–	Yes
–	–	–	–	No

$P(x_4/yes)$

- $P(y=yes/xi) = \dfrac{P(yes) * P(x_1/yes) * P(x_2/yes) + P(x_3/yes) + \dots}{\underbrace{P(x_1) * P(x_2) * P(x_3) * P(x_4)}_{\text{Constant} = \text{Ignore}}}$ #Fixed

- $P(y=no/xi) = \dfrac{P(no) * P(x_1/no) * P(x_2/no) + P(x_3/no) + P(x_4/no)}{\underbrace{P(x_1) * P(x_2) * P(x_3) * P(x_4)}_{\text{Constant} = \text{Ignore}}}$ #fixed

$\boxed{P(yes/xi) = 0.13}$, $\boxed{P(no/xi) = 0.05}$

Normalization ⇓

$> 0.5 \Rightarrow 1$
$< 0.5 \Rightarrow 0$

- $P(yes/xi) = \dfrac{0.13}{0.13 + 0.05} = 0.72 = \boxed{72\%}$ $\longrightarrow$ So, we choose $P(yes/xi)$

- $P(no/xi) = 1 - 0.72 = 0.28 = \boxed{28\%}$

Example: 2 :- Binary classification problem

Day (x_1)	Outlook (x_2)	Temperature (x_3)	Humidity (x_4)	Wind (x_5)	Play Tennis (y)
1	Sunny	hot	high	weak	No
2	Sunny	hot	high	strong	No
3	Overcast	hot	high	weak	Yes
4	Rain	mild	high	weak	Yes
5	Rain	cool	normal	weak	Yes
6	Rain	cool	normal	strong	No
7	Overcast	cool	normal	strong	Yes
8	Sunny	mild	high	weak	No
9	Sunny	cool	normal	weak	Yes
10	Rain	mild	normal	weak	Yes
11	Sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	Rain	mild	high	strong	No

- Outlook (x_1)

	yes	no	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5
Total :-	9	5		

P(sunny|yes) ↓ P(yes)

- Temperature (x_2)

	yes	no	P(yes)	P(no)
Hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cold	3	1	3/9	1/5
Total :	9	5		

- Play (y)

		P(yes)	P(no)
Yes	9	9/14	5/14
no	5		
Total	14		

* if New Data :- Test (Sunny, Hot) ⟹ O/P ?

- $$P(yes/(sunny, HOT)) = \frac{P(yes) * P(Sunny/yes) * P(Hot/yes)}{\cancel{P(Sunny) * P(Hot)}}$$

$$= \frac{9}{14} * \frac{2}{9} * \frac{2}{9}$$

$$= \frac{2}{63} = \boxed{0.031}$$

- $$P(no/(sunny, HOT)) = \frac{P(no) * P(sunny/no) * P(HOT/no)}{\cancel{P(Sunny) * P(HOT)}}$$

$$= \frac{5}{14} * \frac{3}{5} * \frac{2}{5} = \frac{3}{35} = \boxed{0.0857}$$

→ Normalization :-

* $P(yes \mid sunny, hot) = 0.031 = 1 - 0.73 = 0.27 = \boxed{27\%}$

* $P(no / sunny, hot) = 0.085 = \dfrac{0.085}{0.031 + 0.085} = 0.73$

$$= \boxed{73\%} ✓$$

$\boxed{1\ Que}$:- if (Sunny, hot) person play tennis → Yes or NO

Ans:- No ⇒ 73 %

$\boxed{2\ Que}$:- if (overcast, Mild) ⟶ Yes, No?

A:-

- $P(yes / overcast, mild) = \dfrac{P(yes) * P(overcast/yes) * P(mild/yes)}{\cancel{P(overcast) * P(mild)}}$

$$= \frac{9}{14} * \frac{4}{9} * \frac{4}{9}$$

$$= \frac{8}{63} = 0.126$$

- $P(no \mid overcast, mild) = \dfrac{P(no) * P(overcast/no) * P(mild/no)}{\cancel{P(overcast) * P(mild)}}$

$$= \frac{\cancel{5}}{14} * \frac{0}{\cancel{5}} * \frac{2}{5}$$

$$= 0.$$

Normalization

* $P(\text{yes} / \text{overcast}, \text{mild})$ $0.126 = \frac{0.126}{0 + 0.126} = 1$ ✓

$P(\text{no} / \text{overcast}, \text{mild}) = 0 = 1 - 1 = 0$ //

Ans:- Yes ⇒ 100% (if overcast, mild) Play Tennis (or) not.

03/5/22 3:30 Am

CODE :-

* Spam classification.