Machine learning

- **Supervised**
- **unsupervised**

Supervised

Regression:
1. lin reg ✓
2. svm | svr
3. DTR
4. RFR
5. AB | GB | XB.

Classification:
1. Log reg.
2. svc | svm.
3. DTC
4. RFC
5. ABC | GBC | XBC

unsupervised

1. Kmeans.
2. Hirarchical
3. DBScan.

KNN → K-Nearest Negibour.

Naive Bayes → Probability.

Coin = Head → $\frac{1}{2}$

Card = king. → $\frac{4}{52} \Rightarrow \frac{1}{13}$

9 10 11 12 13
[2 - 10] A, J, K, Q.

1. Dimond → 13
2. Heart → 13
3. spade → 13
4. club → 13

$\frac{4}{52} = \frac{1}{13}$

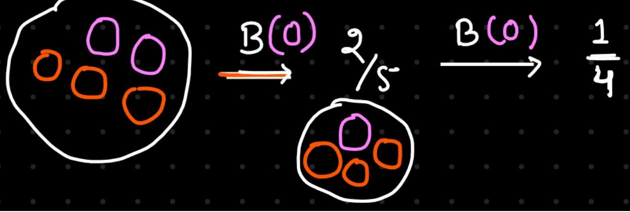
1. independent → each event is not affect by other event
2. Dependent → In dependent event an event affected by another event.
3. mutully exclusive → 1. turning L OR R. 2. Head or tells. 3. King and queen are mutually exclusive — you can't do both at the same time.

1. Coin ⇒ H or T

1. $\frac{1}{2} = 0.5 = 50\%$
2. H or T = 0.5

2. Dependent event eg.

First event — 1 | Second event — 2



52 event 1 (King) $\frac{4}{52}$ event-2 (King) $\frac{3}{51}$

Dependent event.

Diff b/w exclusive and independent event.

ME ⇒ Two event can not occour at the same time [T]

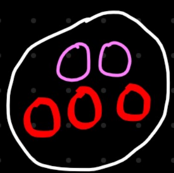
IDE ⇒ One event remains unaffected by the occurance of the other event.

$[T-1] \rightarrow (T)$

independent Coin.

$\frac{1}{2}$ $\frac{1}{2}$

Q. find out the Probability of getting Blue ball.



event (A) event (B)

$P(\text{Blue}) \rightarrow \frac{2}{5}$ $P(\text{Blue}) \rightarrow \frac{1}{4}$ $\Rightarrow \frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10} = 0.1$

$P(A)$ $P(B|A)$ [B is happening and a is already happend]

"event B given event A."

$$\boxed{P(A \text{ and } B) = P(A) \times P(B|A)} \quad - (1)$$

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

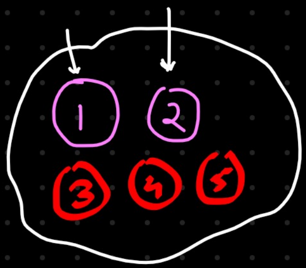
$$\boxed{P(B \text{ and } A) = P(B) \times P(A|B)} \quad - (2)$$

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

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

Bayes theoram.

$$\boxed{P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}} \rightarrow \text{Naive bayes formula.}$$



P(A and B)

$$P(\text{Blue}) \longrightarrow \frac{2}{5} \qquad P(\text{Blue}) \longrightarrow \frac{1}{5} \qquad P(A) \quad P(B|A) \quad = \frac{2}{5} \times \frac{1}{4}$$

$$P(B \text{ and } A) = \frac{2}{5} \times \frac{1}{4}$$

$\Rightarrow$ toss the coin 3 times in a consecutive manner and find out the Prob. of getting head.

$$\boxed{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}}$$

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

$B \to y$, $A \to (x_i)$, $P(B) \to y$, $P(A|B) \to p(x_i|y)$, $P(A) \to P(x_i)$

Dependent feature.

Dataset $\Rightarrow$ $X_1 \quad X_2 \quad X_3 \quad X_4 \quad \text{------} \quad X_n \quad y \rightarrow$ O/P

I/P feature. independent

$$P(y/(x_1, x_2, x_3 \ldots x_n)) = \frac{P(y) \times P(x_1, x_2, x_3, \ldots x_n | y)}{P(x_1, x_2, x_3, x_4, x_5 \ldots x_n)}$$

Data $\longrightarrow$

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

classifier model.

Probability based

Naive bayes.

$\underline{y}$ $(x_1 .. x_n)$ $(x_1 - x_n)$ $\frac{1}{y}$ $P(y)$ $P(x_1 - x_n)$

Senerio.

Binary class Problem.

x_1 x_2 x_3 O/P

yes

NO

yes. $$P(y|x_1 x_2 x_3) = \frac{P(yes) \times P(x_1|yes) \times P(x_2|yes) \times P(x_3|yes)}{\cancel{P(x_1) \times P(x_2) \times P(x_3)}}$$

NO $$P(NO|x_1 x_2 x_3) = \frac{P(NO) \times P(x_1|NO) \times P(x_2|NO) \times P(x_3|NO)}{\cancel{P(x_1) \times P(x_2) \times P(x_3)}}$$

$\frac{\cancel{50}^{5}}{\cancel{10}}$ $\frac{\cancel{100}^{10}}{\cancel{10}}$ 5 10 $\frac{\cancel{50}}{\cancel{100}} = \frac{1}{2}$ $\frac{\cancel{8}}{\cancel{16}} = \frac{1}{2}$

Let's assume. $P(y|x_i) = 0.13$
$P(N|x_i) = 0.05$

$= \frac{0.13}{0.13 + 0.05} = \frac{0.13}{0.18} = 72\%$

$\frac{0.05}{0.13 + 0.05} = \frac{0.05}{0.18} = 28\%$

$P(y|x_i) > P(N|x_1)$

yes

P(Class/data) = (P(data/Class) x P(class))/ p(data)

$$P(y/(x_1, x_2 \dots x_n)) = \frac{P(y) \times P(x_1|y) \times P(x_2|y) \times P(x_3|y) \times P(x_4|y) \dots n}{P(x_1) \times P(x_2) \times P(x_3) \dots P(x_n)}$$

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

(14)

Naive bayes classifier.

$$\boxed{P(y) = \frac{9}{14} \quad P(n) = \frac{5}{14}}$$

Outlook

	yes	NO	P(y)	P(N)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5
	9	5		

Temperature.

	Yes	NO	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
	9	5		

Ques. if [Sunny, HoT] → O/P   N

$$P(yes/[Sunny, HoT]) = \frac{P(yes) * P(Sunny/yes) * P(HoT|yes)}{P(Sunny) * P(HoT)}$$

$$P(NO|[Sunny, HoT]) = \frac{P(NO) * P(Sunny|NO) * P(HoT|NO)}{P(Sunny) * P(HoT)}$$

$$P(yes/[Sunny, HoT]) = \frac{P(yes) * P(Sunny/yes) * P(HoT|yes)}{P(Sunny) * P(HoT)} = \frac{9}{14} \times \frac{2}{9} \times \frac{2}{9} = \frac{2}{63} = 0.031$$

$$P(NO|[Sunny, HoT]) = \frac{P(NO) * P(Sunny|NO) * P(HoT|NO)}{P(Sunny) * P(HoT)} = \frac{5}{14} \times \frac{3}{5} \times \frac{2}{5} = \frac{3}{35}$$

$$= 0.085$$

$$P(yes|[Sunny, HoT]) = \frac{0.031}{0.031 + 0.085} = \frac{0.031}{0.116} = 0.27$$

$$= 27\%$$

$$P(NO \mid (\text{Sunny}, \text{Hot})) = \frac{0.085}{0.031 + 0.085} = \frac{0.085}{0.116} = 0.732$$

$$= 73\%$$

Task: 1. Expand humidity and wind like i did with outlook and temprature then you have to solve the Particular Condition.

$$[\text{Sunny}, \text{Hot}, \text{high}, \text{Strong}] \rightarrow O/P$$

$$P(\text{yes} \mid [\text{Sunny}, \text{Hot}, \text{high}, \text{Strong}]) = ?$$

$$P(\text{No} \mid [\text{Sunny}, \text{hot}, \text{high}, \text{Strong}]) = ?$$

task No. 2 ⇒ https://medium.com/analytics-vidhya/use-naive-bayes-algorithm-for-categorical-and-numerical-data-classification-935d90ab273f