$$P(\text{Spam if "Buy" \& "Cheap"}) = \frac{\frac{20}{25} \times \frac{15}{25} \times \frac{25}{100}}{\frac{20}{25}\,\frac{15}{25}\,\frac{25}{100} + \frac{5}{75}\,\frac{10}{75}\,\frac{75}{100}}$$

$$= 94.737\%$$

Naive Baye's Theorem

* Baye's Theorem

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

given

Dataset

$$x = \{\underbrace{x_1, x_2, x_3 \ldots x_n}_{\text{I/P Feature}}\}\{Y\}$$

o/p lable

$$P(Y|x_1, x_2, \ldots x_n) = \frac{\overbrace{P(x_1|Y)\,P(x_2|Y)\,P(x_3|Y) \ldots P(x_n|Y)}^{P(B|A)} * \overbrace{P(Y)}^{P(A)}}{\underbrace{P(x_1)\,P(x_2)\,P(x_3) \ldots P(x_n)}_{P(B)}}$$

given

$$= \frac{P(Y) \prod_{i=1}^{n} P(x_i|Y)}{P(x_1)P(x_2)P(x_3) \ldots P(x(n))} \leftarrow \text{constent}$$

$$P(Y|x_1, x_2, \ldots x_n) \propto P(Y) \prod_{i=1}^{n} P(x_i|Y)$$

$$Y = \text{argmax}_Y\, P(Y) \prod_{i=1}^{n} P(x_i|Y)$$

Yes	9	9/14
No	5	5/14
Tot	14	100%

Outlook

	Yes	No	P(Y)	P(N)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

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
Total	9	5	100%	100%

Today (SUNNY, HOT)

$$P(\text{Yes}|\text{Today}) = \frac{P(\text{SUNNY}/\text{Yes}) \times P(\text{HOT}/\text{Yes}) * P(\text{Yes})}{P(\text{Today})}$$

$$= \frac{2}{9} \times \frac{2}{9} \times \frac{9}{14} = 0.031$$

Page :
Date : / /

$P(\text{No}|\text{Today}) = \frac{3}{5} \times \frac{2}{5} \times \frac{5}{14} = 0.08571$

Normalise $p(\text{yes}) = \frac{0.031}{0.31 + 0.08571} = 0.27$

$P(N) = 1 - 0.27 = 0.73$ — ~~Play~~

So Today (SUNNY, HOT) — NO

Q) Consider the given Datset, apply Naive-Baye's algorithm & predict that if a fruit has the following properties then which type of fruit it is.

Fruit = { Yellow, Sweet, Long } = X

Fruit	Yellow	Sweet	Long	Total
mango	350	450	0	650
Banana	400	300	350	400
others	50	100	50	150
Total	800	850	400	1200

Naiv Baye's

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

Probability of A when B is true.

i) Assume fruit is mango

$P(X|\text{mango})$

$$P(\text{Yellow}|\text{mango}) = \frac{P(\text{mango}|\text{Yellow})\,P(\text{Yellow})}{P(\text{mango})}$$

$$= \frac{\left(\frac{350}{800}\right) \cdot \left(\frac{800}{1200}\right)}{\left(\frac{650}{1200}\right)} = \frac{350}{650} = 0.53$$

$P(\text{Sweet}|\text{mango}) = 0.69$

$P(\text{Long}|\text{mango}) = 0$

$\therefore P(X|\text{mango}) = P(Y|M)\,P(S|M)\,P(L|M) = 0$

ii) Assume fruit is Banana.

$$P(\text{Yellow}|\text{Banana}) = \frac{P(\text{Banana}|\text{yellow})\,P(\text{Yellow})}{P(\text{Banana})} = \frac{\left(\frac{400}{800}\right)\left(\frac{800}{1200}\right)}{\left(\frac{400}{1200}\right)} = 1$$

$P(S|B) = 0.75$ $\quad P(L|B) = 0.875$

$\therefore P(X|\text{Banana}) = 1 \times 0.75 \times 0.875 = \underline{\underline{0.65}}$.

iii) ~~oth~~ Fruit ~~o~~ otheres

$$P(Y|\text{others}) = \frac{P(\text{others}|\text{Yellow})\,P(\text{Yellow})}{P(\text{others})} = \frac{\left(\frac{50}{800}\right)\left(\frac{8}{12}\right)}{\left(\frac{150}{1200}\right)} =$$

$$P(L|\text{others}) = \frac{P(\text{others}|\text{Long})\,P(\text{Long})}{P(\text{others})} = \frac{\left(\frac{50}{400}\right)\left(\frac{400}{1200}\right)}{(150/1200)} =$$

$$P(S|\text{others}) = \frac{P(\text{others}|S)\cdot P(\cancel{\text{others}}\ \text{Sweet})}{P(\text{others})} = \frac{\left(\frac{100}{850}\right)\cdot\left(\frac{850}{1200}\right)}{\left(\frac{150}{1200}\right)} =$$

$$P(X|\text{others}) = 0.33 \times 0.66 \times 0.33 = \underline{0.072}$$

$\therefore P(X|M) = 0 \qquad P(X|B) = 0.65 \qquad P(X|O) = 0.072$

$\therefore$ Fruit is Banana.

Q.2) For the given dataset, Apply Naive-Baye's Algorithm & Predict the outcome for a Car = {Red, Domestic, SUV}

X = [Red, Domestic, SUV]

Color	Type	Origin	Stolen
Red	Sport	Domestic	Y
Red	Sport	D	~~Y~~ No
Red	Sport	D	Yes
Yellow	Sport	D	No
Yellow	Sport	Imported	Yes
Yellow	SUV	I	No
Yellow	SUV	I	Yes
Yellow	SUV	Domestic	No
Red	SUV	I	No
Red	Sport	I	Yes

$$\underbrace{P(A|B)}_{\text{Posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}}\,P(A) \leftarrow \text{Proposition}}{P(B) \leftarrow \text{Evidance}}$$

Page :
Date : / /

i) $P(Red|Yes) = \frac{P(Yes|Red)\,P(\cancel{Yes}\ Red)}{P(Yes)} = \frac{(\frac{3}{5})(\frac{5}{10})}{(5/10)} = \frac{3}{5}$

ii) $P(D|Yes) = \frac{P(Yes|D)\,P(D)}{P(Yes)} = \frac{(2/5)\,(5/10)}{(5/10)} = \frac{2}{5}$

iii) $P(S^{UV}|Yes) = \frac{P(Yes|S)\,P(S)}{P(Yes)} = \frac{(1/5)(5/10)}{(5/10)} = \frac{1}{5}$

$P(X|Yes) = \frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} = \frac{\cancel{24}\ 6}{125} = \cancel{0.192}\ 0.024$

i) $P(Red|N) = \frac{P(N|R)\,P(R)}{P(N)} = \frac{(2/5)(5/10)}{(5/10)} = \frac{2}{5}$

ii) $P(D|N) = \frac{P(N|D)\,P(D)}{P(N)} = \frac{(3/5)\,(5/10)}{(5/10)} = 3/5$

iii) $P(SUV|N) = \frac{P(\cancel{SUV}|SUV)\,P(SUV)}{P(N)} = 2/5$

$P(X|N) = \frac{2}{5} \times \frac{3}{5} \times \frac{2}{5} = \frac{12}{125} = 0.072$

$\therefore P(X|Yes) = 0.024 \qquad P(X|N) = 0.072$

$\therefore$ X [Red, Domestic, SUV) Not stolen.

Naive Bayes

i) ML algorithm . Supervised Learning
ii) Used as a classifier for texts
iii) Based on Bayes Theorem

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

Likelihood . prior probability
posterior

2) ~~Is Student~~ buy ~~computer~~ ?
It says the posterior probability of a hypothesis given the data is given by probability

$$P(h|D) = \frac{P(D|h)\,P(h)}{P(D)}$$

What is Naive Bayes Algorithm

1) Naive Bays algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/ class.
2) For instance, if your are trying to identify a fruit based on its colour, shape & taste, then an orage colored, spherical & ~~tangy~~ tangy fruit would most likely be an oragne.
3) All of these probability individually contribute to the probability that this fruit is an orage & ie why it is know as 'naive'.
4)
P(A|B) :- Probability (conditional probability) of occurance of event A given the event B is true
P(A) P(B) :- probabilities of occurance of event A & B
P(B|A) :- Probability of the occurance of event B given the event A is true

1) A is called the proposition & B is called the evidence

2) P(A) is called the prior probability of proposition & P(B) is called prior probability of evidance.

3) P(A|B) is called posterior

4) P(B|A) is the likelihood.

$$\text{Posterior} = \frac{(\text{Likelihood})(\text{Proposition Prior Probability})}{(\text{Evidance prior probability})}$$

* Bayes' Theorem for Naive Bayes Algorithm

In a machine learning classification problem there are multiple features & classes, say The main aim in the Naive Bayes algorithm is to calculate the conditional probability of an object with a feature vector belongs to a particular class,

$$P(C_i|x_1, x_2, \ldots x_n) = \frac{P(x_1, x_2, \ldots x_n | C_i) \cdot P(C_i)}{P(x_1, x_2, \ldots x_n)}$$

$$P(C_i|X) = P(X|C_i)$$

$$P(X|C_i) = \prod_{k=1}^{c} P(X_k|C_i) = P(x_i|C_1) \times P(x_1|C_2) \times \ldots$$

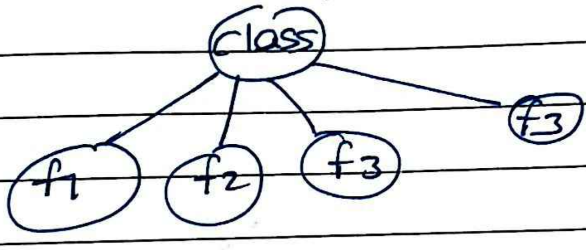
* Naive Bayes uses a similar method to predict the probability of different class based on various attributes.

* This algorithm is mostly used in text classification & with problem having multiple class.

Page :
Date : / /

* The Naive Bayesian Classifier is based on Bayes' theorem with independance assumption bet^n predictors.
* A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which make it particularly useful for very large datasets.

$$P_r[\text{class}/\text{observations}] = \frac{P_r[\text{observation}/\text{class}]\,P_r[\text{class}]}{P_r[\text{observation}]}$$



* Advantage of Naive Bayes Classifier
1) Not sensitive to irrelevant features.
2) Very simple & easy to implement
3) Needs less training data
4) Highly scalable with no. of predictors & data points.
5) As it it is fast it can be used in real time predication
6) Handles both contineous & discrete data

Date : / /

Shopping Demo :- Problem statement.

To predict wheather a person will purchase a product on a specific combination of Day Discount & free Delivery using Naive Bayes Classifier

Day	Discount	Free Delivery		Buy
weekdays	Yes	Yes	probabilities	Don't Buy
weekend	No	No		
Holidays				

Day	Discount	Free Delivery	Purchase
weekday	Y	Y	Y
weekday	Y	Y	Y
weekday	N	N	N
Holiday	Y	Y	Y
weekend	Y	Y	Y
H	N	N	N
weekend	Y	N	Y
weekday	Y	Y	Y
weekend	Y	Y	Y
H	Y	Y	Y
H	N	Y	Y
H	N	N	N
weekend	Y	Y	Y
H	Y	Y	Y

frequency Table.

Frequency Table		Buy Y	Buy N
Discount	Y	19	1
	N	5	5

Frequency Table		Buy Yes	Buy No
Day	weekday	4	2
	weekend	7	1
	Holiday	8	3

Frequency Tabel		Buy Y	Buy N
free D	Y	21	2
	N	3	4

Q) Apply Naive Bayes for given document

Text	Category
A grate game	Sports
The election was over	Nonsports
very clean match	Sports
A clean but forgetable game	Sports
I was a close election	Non Sports

Total words in sport → 11
Total words in nonsports → 9
Total unique words → 14

Classify → "A very close game → Category < Sports / Nonsports

P(Sports / "A very close game") = ?
P(Nonsports / "A very close game") = ?
} whover wins is answer

How to find probability.
Step-1 Feature Engineering ↓

P(A very close game) = P(A) × P(very) × P(close) × P(game)

P(A very close game | sports) = P(A|sports) × P(very|sports) × P(close|sport) × P(Game|sport)

P(A very close game | sports) = P(A|Nonsports) × P(very|Nonsports) × ......

Page :
Date : / /

Step 2 : Probability

"A very close game"

$P(a|sports) = \frac{2}{11}$ $P(very|sports) = \frac{1}{11}$ $P(close|sports) = 0/11$

4 hence $P(\text{A very close game}|sports) = 0$.

Solⁿ Laplace Smoothing

$\theta_i = \frac{x_i + \alpha}{N + \alpha d}$ $\alpha > 0$ mainly $\alpha = 1$ (always)

d = No. of unique word or distant words

$\theta_i [P(word)] = \frac{x_i \text{(word count)} + 1}{\text{total No of words} + \text{No. of Unique words)}}$

$P(close|sports) = \frac{0+1}{11+14}$ $P(very|sports) = \frac{1+1}{11+14}$

$P(a|sports) = \frac{2+1}{11+14}$ $P(game|sports) = \frac{2+1}{11+14}$

~~word~~ $P(close|Nonsport) = \frac{1+1}{9+14}$ $P(a|Non sport) = \frac{1+1}{9+14}$

$P(very|Nonsport) = \frac{0+1}{9+14}$ $P(game|Nonsport) = \frac{0+1}{9+14}$

$P(\text{a very close geme}|Sports) = 4.61 \times 10^{-5} = 0.0000461$

$P(\text{a very close game}|Nosports) = 1.43 \times 10^{-5} = 0.0000143$

$P(\text{a very close game}|Sports)$ is high hence it belongs to sports

Naive Bayes

1) So to handle this problem in Naive Bayes make assumption that individual X_i's are independent given Y.

$$\therefore P(Y|X) = P(X_1|Y)\,P(X_2|X_1 Y)\,P(X_n|X_1 \cdots X_{n-1} Y)\,P(Y)$$

$$= P(X_1|Y)\,P_2(X_2|Y)\,P(X_n|Y) \cdot P(Y)$$

So we are assuming conditional independence among the individual attributes $X_1, X_2 \ldots X_n$ & based on this we can do classification.

Page:
Date: / /

* So assuming all the input features are conditionally independent & this can be computed as probability x given Y

Naive Bayes

Bayes Rule $P(Y=y_k \mid x_1 \dots x_n) = \dfrac{P(Y=y_k)\,P(x_1 \dots x_n \mid Y=y_k)}{\sum_j P(Y=y_j)\,P(x_1 \dots x_n \mid Y=y_j)}$

Assuming Conditional independence among x_i's

$$P(Y=y_k \mid x_1 \dots x_n) = \frac{P(Y=y_k)\prod_i P(x_i \mid Y=y_k)}{\sum_j P(Y=y_j)\prod_i P(x_i \mid Y=y_j)}$$

So classification Rule for $X^{new} = \langle x_1 \dots x_n \rangle$ is

$$Y^{new} \leftarrow \arg\max_{y_k} P(Y=y_k)\prod_i P(x_i^{new} \mid Y=y_k)$$

∴ classifier is

Train Naive Bayes

for each value y_k

estimate $\pi_k \equiv P(Y=y_k)$

for each value x_{ij} of each attribute x_i

estimate $\theta_{ijk} \equiv P(x_i = x_{ij} \mid Y=y_k)$

Clasify (X^{new})

$$Y^{new} \leftarrow \arg\max_{y_k} P(Y \equiv y_k)\prod_i P(x_i^{new} \mid Y) = y_k)$$

$$Y^{new} \leftarrow \arg\max_{y_k} \pi_k \prod_i \theta_{ijk}$$

Probabilies must sum to 1 so need estimate only n-1 param

Learning Phase

Outlook	Play=Y	Play=N
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temp	Play=Yes	Play=
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Humidity	Play=Yes	Play=No
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=Yes	Play=N
Strong	3/9	3/5
Weak	6/9	2/5

$$P(\text{Play}|\text{Yes}) = \frac{9}{14} \quad P(\text{Play}=\text{No}) = \frac{5}{14}$$

Test Phase

Given new instance predict its label

$x^1 = (\text{Outlook}=\text{Sunny}, \text{Temp}=\text{Cool}, \text{Humidity}=\text{High}, \text{Wind}=$

$P(\text{Outlook}=\text{Sunny}|\text{Play}=\text{Yes}) = 2/9$

$P(\text{Outlook}=\text{Sunny}|\text{Play}=\text{No}) = 3/5$

Estimate Parameters Y, X_i discrete-valued

$$\hat{\pi}_k = \hat{P}(Y=y_k) = \frac{\#D\{Y=y_k\}}{|D|}$$

$$\theta_{ijk} = \hat{P}(X_i=x_{ij}|Y=y_k) = \frac{\#D\{X_i=x_{ij} \wedge Y=y_k\}}{\#D\{Y=y_k\}}$$

MAP estimates

$$\hat{\pi}_k = \hat{P}(Y=y_k) = \frac{\#D\{Y=y_k\}+1}{|D|+1R}$$

$$\theta_{ijk} = \hat{P}(X_i=x_{ij}|Y=y_k) = \frac{\#D\{X_i=x_{ij} \wedge Y)=y_k\}+1}{\#D\{Y=y_k\}+1M}$$

Date : / /

Gaussian Naïve Bayes (Continuous X)

1) Continuous-valued features
Conditional probability modeled with normal distribution

$$P(X_i = x \mid Y = y_k) = \frac{1}{\sigma_{ik}\sqrt{2\pi}} e^{-\frac{(x-\mu_{ik})^2}{2\sigma_{ik}^2}}$$

Assume variance - is independent of Y (ie, σ_i)
independent