Naive Bayes

Naive Bayes → Classification → Textual Data

Textual Data → Senitemal Analysis

Textual Data → Spam filter

Naive Bayes → Baseline Mode

→ XgBoost

→ Deeplearning Model

Intuition

Outlook	Temperature	Humidity	Windy	Play Tennis
Sunny	Hot	High	false	No
Sunny	Hot	High	True	No
Overcast	Hot	High	false	Yes
Rainy	Mild	High	false	Yes
Rainy	cool	Normal	false	Yes
Rainy	cool	Normal	True	No.
Overcast	cool	Normal	True	Yes
Sunny	Mild	High	false	No
Sunny	cool	Normal	false	Yes
Rainy	Mild	Normal	false	Yes
Sunny	Mild	Normal	True	Yes
overcast	Mild	High	True	Yes
Overcast	Hot	Normal	false	Yes
Rainy	Mild	High	True	No

$\boxed{\text{Naive Bayes}}$

$\boxed{\text{Bayes theorem}}$

$\{$ Sunny, cool, Normal, True $\} = W$

↑

query point

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

Y/S

2 probability

Yes: 0.82

No: 0.18

$$P(Yes/W) = P(W/Yes)\, \overset{9/14}{P(Yes)}$$

$$P(No/W) = P(W/No)\, \overset{5/14}{P(No)}$$

$0 = \boxed{P(Yes/W)}$ ↳ $\boxed{P(sunny \cap cool \cap Normal \cap true/Yes)}$

0 combination of this

Yes in data = 9

No in data = 5

$0 = \boxed{P(No/W)}$ ↳ $\boxed{P(sunny \cap cool \cap Normal \cap true/No)}$

In 5 rows of 'NO' there's 0 combination of ↑

* Because 0 combination of both Probability, So Naive Bayes take Naive assumption.

$\boxed{P(sunny \cap cool \cap Normal \cap true/Yes)}$

↳ $\underset{2/9}{P(sunny/Yes)}\ \underset{3/9}{P(cool/Yes)}\ \underset{6/9}{P(Normal/Yes)}\ \underset{3/9}{P(True/Yes)}$

$(2/9) \times (3/9) \times (6/9) \times (3/9) \times (9/14) \rightarrow$ Yes/Total

$\boxed{P(\text{sunny} \cap \text{cool} \cap \text{Normal} \cap \text{true} / \text{No})}$

↳ $P(\text{sunny}/\text{No}) \quad P(\text{Cool}/\text{No}) \quad P(\text{Normal}/\text{No}) \quad P(\text{True}/\text{No})$

$\frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{3}{5} \times \frac{5}{14}$

* After multiply which one have high Prob is the answer.

Mathematical formulation

$x_1 \quad x_2 \; ------ \; x_n \qquad Y$ (multiclass classification) $1, 2, 3, ------ k$

k classes

$\langle x_1, x_2, x_3 ------ x_n \rangle \rightarrow$ Naive bayes $\qquad$ k prob

$$\begin{cases} P(Y_1/X_T) = \dfrac{P(X_T/Y_1)\,P(Y_1)}{P(X_T)} \\[2ex] P(Y_2/X_T) = \dfrac{P(X_T/Y_2)\,P(Y_2)}{P(X_T)} \\ \vdots \\ P(Y_K/X_T) = \dfrac{P(X_T/Y_2)\,P(Y_K)}{P(X_T)} \end{cases}$$

$$X_T = \langle x_1, x_2, ------ x_n \rangle$$

$$P(Y_K/X_T) = P(X_T/Y_K)\,P(Y_K)$$

$$= P(x_1 \cap x_2 \cap x_3 \cap ------ \cap x_n / y_k)\,P(Y_K)$$

$\cap = ,$

$$= P(x_1, x_2, x_3, ------, x_n / Y_K)\,P(Y_K)$$

$$= P(\overbrace{\boxed{x_1, x_2, x_3, \ldots, x_n}}^{A} / \overbrace{\boxed{Y_k}}^{B})\, P(Y_k)$$

$$= \frac{P(x_1, x_2, x_3, \ldots x_n, Y_k)\, \cancel{P(Y_k)}}{\cancel{P(Y_k)}} \qquad P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$= P(\overbrace{\boxed{x_1}}^{A}, \overbrace{\boxed{x_2, x_3, \ldots x_n, Y_k}}^{B})$$

$$= P(x_1 / x_2, x_3, \ldots x_n, Y_k)\; P(\overbrace{\boxed{x_2}}^{A}, \overbrace{\boxed{x_3, \ldots, x_n, Y_k}}^{B})$$

$$P(x_2 | x_3, x_4, \ldots x_n, Y_k)\; P(\overbrace{\boxed{x_3}}^{A}, \overbrace{\boxed{x_4, \ldots x_n, Y_k}}^{B})$$

$$= P(x_1 | x_2, x_3 \ldots x_n, Y_k)\; P(x_2 | x_3, x_4 \ldots Y_k) \ldots$$

$$\ldots P(x_{n-1} | x_n, Y_k)\; P(x_n / Y_k)\; P(Y_k)$$

<u>let</u> <u>Naive</u> <u>Assumption</u> → feature are independent of each other but dependent on output.

x_1	x_2	x_3	----	x_n

idependent

$$\boxed{P(A/B) = P(A)}$$

$$\boxed{P(A / B \cap C) = P(A / B, C) = P(A/C)}$$

$$Y_1 = P(x_1 | Y_k)\; P(x_2 | Y_k)\; P(x_3 | Y_k) \ldots P(x_{n-1} / Y_k)\; P(x_n | Y_k)\; P(Y_k)$$

dependent

$\vdots$

Y_k

How Naive Bayes handles numerical data?

age	married
27	Y
61	N
52	Y
-	-
-	-
-	-

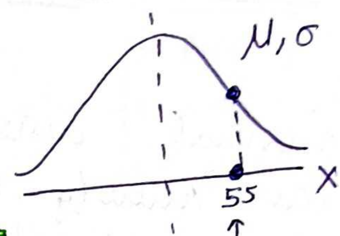
$\{55\} \rightarrow Y, N?$

query

$P(Y/55) = \underline{P(55/Y)}\ P(Y)$ → 0 , → 0.62

$P(N/55) = \underline{P(55/N)}\ P(N)$ → 0

* if 55 is not present in data then probability is 0



σ, μ (find)

let $\boxed{\text{data}}$ is $\boxed{\text{gausian distribution}}$

'Y' category data and 'N' category data individual

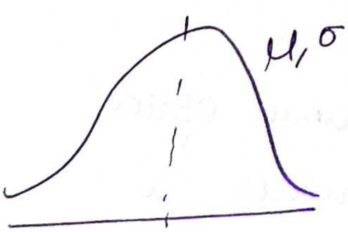
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \boxed{0.62}$$

(x → 55)

$f(x)$ → Probability density

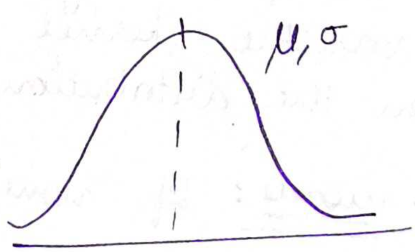
* In Married column also assume that $\boxed{\text{'Y' category}}$ is gausian distribution and find σ, μ. And $\boxed{\text{'N' category}}$ is also gausian distributed and also find σ, μ seprately.

'Y' category



gaussian distribution

'N' category



What is data is not Gaussian?

1. Data Transformation: Depending on the nature of your data, you could apply a transformation to make it normally distributed. Common transformation include the logarithm, square root, and reciprocal transformation.

2. Alternative Distribution: If you know or suspect that your data follow a specific non-normal distribution (e.g exponential, Poisson, etc). you can modify the Naive Bayes algorithm to assume that specific distribution when calculating the likelihoods.

3. Discretization: You can turn your continuous data into categorical data by binning the values. There are various ways to decide on the bins, including equal width bins, equal frequency bins, or using a more sophisticated method like k-means clustering. Once your data is binned, you can use the standard Multinomial or Bernoulli Naive Bayes methods.

4. Kernel Density Estimation: A non-parametric way to estimate the probability density function of a random variable. Kernel density estimation can be used when the distribution is unknown.

5. Use other models: If none of the above option works well, it may be best to consider a different classification algorithm that doesn't

make strong assumption about the distribution of the feature, such as Decision Trees, Random Forests, or Support Vector Machine.

Naive Bayes on Text Data

Sentiment Analysis

Review	Sentiment
	+
	−
	+

1) Text preprocessing

2) Vectorize → text → number
   - → BOW
   - → OHE
   - → Embeddings

3) apply naive bayes

BOW

→ 10L words → 35000 uniques

↳ 5000 most frequently used words

Reviews	Sentiment
I like the movie	+ve
really hated the movie	−ve

hate	like	movie	actor	---
0	1	1	0	---
2	0	3	2	---
3	2	1	1	---
⋮				

(50000, 50000)

↳ shaped

Adj - - - great movie

numbers

[0   2   1 - - - - 5] → (15000)

Vector

+ve   -ve

$P(+ve \mid hate=0,\ like=2,\ \text{------}) = \rightarrow 0.37$

$P(-ve \mid hate=0,\ like=2,\ \text{- ---}) = \rightarrow 0.23$

→ +ve

positive Sentiment

$P(hate=0 \mid +ve)\quad P(like=2 \mid +ve)$ ------

$P(hate=0 \mid -ve)\quad P(like=2 \mid -ve)$ ---

Numerical Stability

Underflow → represent decimal in memory
floating point

Advance Topic ↓

What is underflow in computing?

Underflow is a condition that can occur in computing when a number nears zero and the computer can no longer store it accurately in memory using floating-point representation. It happens when a calculated result is a smaller absolute value than the computer can actually represent

Most computers use a form of representation called floating-point to represent real numbers. This representation has a certain precision limit, and it can only represent numbers between a certain minimum and maximum value. If a number is too close to zero (but not zero), it might be smaller than the smallest representable positive number in the machine's floating point representation. When an operation on such small number is performance, the machine might round the result to zero, leading to a loss of precision.

Underflow can be a problem in certain domains, such as machine learning, where calculations often involve probabilities. Probabilities are positive numbers that can be very close to zero. When multiplying many small probabilities together, the result can underflow. One common way to avoid underflow in such scenarios is to perform calculation in the log domain, where addition and subtraction are used instead of multiplication and division, thereby maintaining higher numerical precision.

cgpa | iq | placement $\{8.1, 81\} \rightarrow Y/N$

$$P(Y/8.1, 81) = P(Y)\ P(8.1/Y)\ P(81/Y)$$

→ 0 probability

$$P(N/8.1, 81) = \text{''}$$

→ 0 probability

let if probability of multiplication is = $0.1 \times 0.3 \times 0.4 \times 0.6 \ldots$

$= 0.000046$

very small number and round off is = 0 and showing underflow.

Solution

Log probabilities

$$\log(P(a)\ P(b)\ P(c) \ldots)$$

$$= \log(P(Y)) + \log(P(8.1/Y)) + \log(P(81/Y)) \ldots$$

$\therefore [\log(ab) = \log a + \log b]$

$$= \log(P(N)) + \log(P(8.1/N)) + \log(P(81/N)) \ldots$$

* which one have large probability is the Output.

Laplace Additive Smoothing

Review			Sentiment
W_1	W_2	W_3	0
W_1	W_3	W_3	1
W_2	W_2	W_1	0

$\rightarrow$

Review	W_1	W_2	W_3	Sentiment
s_1	1	1	1	0 −ve
s_2	1	0	1	1 +ve
s_3	1	1	0	0 −ve

$\boxed{s_4} \rightarrow \boxed{W_1 \quad W_1 \quad W_1} \rightarrow$ +ve, −ve?

1 0 0

$$\underline{P(+ve/s_4)} = P(+ve)\ P(W_1=1/+ve)\ P(W_2=0/+ve)\ P(W_3=0/+ve)$$

$$\boxed{\left(\tfrac{1}{3}\right)} \times \left(\tfrac{1}{1}\right) \times \left(\tfrac{1}{1}\right) \times \left(\tfrac{0}{1}\right) = 0$$

$$\underline{P(-ve/s_4)} = P(-ve)\ P(W_1=1/-ve)\ P(W_2=0/-ve)\ P(W_3=0/-ve)$$

$$\left(\tfrac{2}{3}\right) \times \left(\tfrac{2}{3}\right) \times (0) \times \left(\tfrac{1}{2}\right) = 0$$

* here log will not use cuz log(0) = undifined

Solution: add some value at numinator (α) and denominator ($n\alpha$)

$$\frac{1+\alpha}{3+n\alpha}$$

$\boxed{\alpha=1} \rightarrow$ default

$\boxed{n=2} \rightarrow$ depend on (~~Gaussian~~, bernauli, multinomial)

$n \rightarrow$ vary $\rightarrow$ depend

$$\frac{1+1}{3+2(1)} = \frac{2}{5}$$

$$\frac{0}{1} = \frac{0+1}{1+2(1)} = \frac{1}{3}$$

Probability never be 0.

Bias Variance Tradeoff

$\log(P(0))$ $= \boxed{\dfrac{+\alpha}{+n\alpha}}$

↳ Laplace Additive Smoothing

↓

flexibility = control biance and variance

$\boxed{\text{high bias}}$

$\alpha \xrightarrow{\text{change value}}$ low bias

$\boxed{\text{high Variance}}$

$\alpha \xrightarrow{\text{change value}}$ low variance

Intitution

$f_1 \mid f_2 \mid f_3 \mid \text{---} \mid Y$ yes → 500
No → 500

↕ 500 rows

$P(Y/X) = P(Y)\ P(f_1/Y)\ P(f_2/Y)\ P(f_3/Y) \text{---}$

$\frac{0}{500}$ → $\boxed{\alpha = 0,0001}$ ⟹ $\frac{3}{500}$

↳ $\frac{12}{500}$

$\boxed{\text{high Variance}}$

↳ when use small 'α'

let use 'α' → high $= 10000$

$n = 2$

$$\frac{1 + 10000}{500 + 2000} = \frac{10001}{20500} = \text{almost } = \left(\frac{1}{2}\right)$$

$$\frac{2}{500} + \frac{1000}{2000} = \frac{10002}{20500} = \text{almost} = \frac{1}{2}$$

$P(Y/x) = \boxed{P(Y)\ P(f_1/Y)\ P(f_2/Y)\ \cdots\ P(f_n/Y)}$

$P(N/x) = \boxed{P(N)\ P(f_1/N)\ P(f_2/N)\ \cdots\ P(f_n/N)}$

Same output

So it is depend on Number of 'Yes' and 'Nos' present in dataset.

high bias

$\alpha \uparrow$ high ~~bias~~
lead to high bias (underfitting)

$\alpha \downarrow$ low $\simeq 0$
high variance
(overfitting)

Gaussian Naive Bayes

Data $\rightarrow$ all feature are numerical

cgpa | iq | placement (1000 student)

cgpa $\downarrow$ num, iq $\downarrow$ num

1000 student $\rightarrow$ 500Y, 500N

500: cgpa | iq | placement — Y, Y, Y, Y, ... $\rightarrow$ Normally distributed (Gaussian)

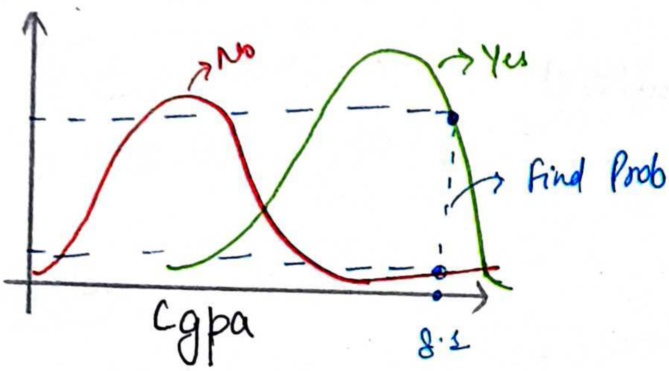
500: cgpa | iq | place — N, N, N, N, ... $\rightarrow$ Normally distributed (Gaussian)

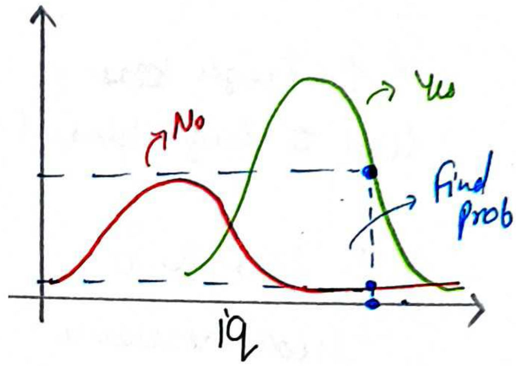
{8.1, 81} → Y/N?

$P(Y / cgpa = 8.1, iq = 81) = \overset{\nearrow prior}{P(Y)} \; \underline{P(cgpa = 8.1 / Y)} \; \underline{P(iq = 81 / Y)}$

$P(N / cgpa = 8.1, iq = 81) = P(N) \; \underline{P(cgpa = 8.1 / N)} \; \underline{P(cgpa = 81 / Y)}$

* Some time Probability will be 0. Because exact 8.1 or 81 not present in data. So we use Gaussian





Categorical Naive Bayes

Categorical Naive Bayes is a variant of the Naive Bayes algorithm designed specifically to handle categorical data.

$\boxed{\text{Data}}$ → all feature are categorical

x_q {Sunny, Hot, High False} → Y/N?

$P(Y/x_q) = \underset{9/14}{P(Y)} \; \boxed{\overset{\nearrow likelihood}{P(Out = Sunny / Y)} \; P(Temp = Hot / Y) \ldots}$

$P(N/x_q) = \underset{5/14}{P(N)} \; \boxed{P(Out = Sunny / N) \; P(temp = Hot / N)}$

Laplace Additive smoothing

$$P(\text{Out} = \text{sunny}/N) = \frac{\overset{n_{tc}}{3} + \alpha}{\underset{n_c}{5} + \alpha(n_i)} = \frac{3+1}{5+1(3)}$$

n_i → Outlook column → 3 categories

$$P(\text{temp} = \text{Hot}/N) = \frac{2+1}{5+1(3)}$$

→ Temperature column → 3 categories

Outlook	Temperature	Humanity	Windy	Play Tennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Multinomial Naive Bayes

Multinomial Naive Bayes is a variant of the Naive Bayes algorithm that is particularly suited for classification tasks involving discrete features, such as text classification where features correspond to word counts or frequency within the documents.

'P(w/c) = (T_{c,w} + 1) / (text_c + B)'

Here, 'T_{c,w}' is the count of word 'w' in class 'c', 'text_c' is the total count of word in class 'c', and 'B' is the size of the vocabulary.

Multinomial Naive Bayes — discrete

f1	f2	f3	f4	Y
2	1	4	3	0
1	2	10	11	1
0	5	6	11	1
9	1	2	3	0

movie-review	Sentiment
great - - -	1
— — —	0
— —	1

BOW → binary, non-binary discrete

great	epic	like	Y
2(1)	0(0)	0(0)	-
0(0)	3(1)	2(1)	-
1(1)	1(1)	1(1)	-

Multinomial Naive Bayes → count bow, Fractions → Textual data

docID	words in document	in c = China?
1	Chinese Bejing Chinese	Yes
2	Chinese Chinese Shanghai	Yes
3	Chinese Macao	Yes
4	Tokyo Japan Chinese	No
5	Chinese Chinese Chinese Tokyo Japan	?

Training set ← 1–4

test set ← 5

↓ Terminal BoW

	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan	
d1	2	1	0	0	0	0	Y
d2	2	0	1	0	0	0	Y
d3	1	0	0	1	0	0	Y
d4	1	0	0	0	1	1	N

test set →

P(Y / Chin = 3, bej = 0, shan = 0, Mac = 0, Toc = 1, Jap = 1)

P(N / Chin = 3, bej = 0, shan = 0, Mac = 0, Toc = 1, Jap = 1)

P(Y) → 3/4 P(Chinese/Y) P(beijing/Y) ----- P(Japan/Y)

P(Chinese/Y) → $\frac{5}{8}$

P(beijing/Y) → $\frac{1}{8}$

P(Japan/Y) → $\frac{0}{8}$

8 → Total word in 'Yes'

* because of '0' in probability, we use laplace Additive Smoothing

$$P(\text{Japan}/y) = \frac{0 + \alpha}{8 + n\alpha}$$

$\boxed{\alpha = 1}$ default

n ↳ $\boxed{\text{size of vocabulary}}$

↳ Unique Value eg:- Chinese, beijing, Japan - - -

$$\frac{0+1}{8+(6\times 1)} = \frac{1}{14}$$

$$P(\text{Chinese}/N) = \frac{1+1}{3+(6\times 1)} = \frac{2}{9}$$

3 ↳ Total word in 'No'

2nd step ↓

$(P(\text{Chinese}/Y))^{3}$ → 3 Chinese present in test set

$(P(\text{beijing}/Y))^{0}$ → 0 beijing present in test set

↳ Ans = 1

⋮

$(P(\text{Japan}/Y))^{1}$ → 1 japan present in test set

$P(Y)$

$$= \frac{3}{4} \times \left(\frac{3}{7}\right)^3 \times \frac{1}{14} \times \frac{1}{14} \approx 0.0003$$

Chinese, tokyo, Japan

* beijing of power is 0 and ans is 1. when multiply with other is same of 1.

* We do same step for No date

$$\overset{\to P(N)}{=} \frac{1}{4} \times \underbrace{\left(\frac{2}{9}\right)^3}_{\text{Chines}} \times \underbrace{\frac{2}{9}}_{\text{Tokyo}} \times \underbrace{\frac{2}{9}}_{\text{Japan}} \approx 0.0001$$

Probability Distributions

An ~~eg~~ engineering college has a placement rate of 0.3 (placed), meaning that any given student has a 0.3 chance of getting placed through campus recruitment. If you randomly select a student, what is the probability that the student is :

1. placed ? → $P(\text{placed}) = 0.3$
2. Not placed ? → $P(\text{not placed}) = 0.7$

↳ Bernaulli → $P(x) \to \{0 \text{ or } 1\}$ ↳ outcome binary

↳ 1 trial

$$P(X=k) = pk + (1-p)(1-k)$$

p = prob of getting

$k = 0, 1$

if $k = 0$

$= (1-p)(1-k)$

if $k = 1$

$= pk$

An engineering college has a placement rate of 0.3, meaning that any given student has a 0.3 chance of getting placed through campus recruitment. If you randomly select 10 students what is the probability that:

1. 9 out of 10 students get placed!

no. of trial = 10

$$\rightarrow \begin{array}{cccccccc} P & P & P & P & \cdots & P & N \\ \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow & \downarrow \\ 0.3 & 0.3 & 0.3 & 0.3 & & 0.3 & 0.7 \end{array} \Rightarrow (0.3)^9 (0.7)^1$$

N P P P P $\Rightarrow$ $\boxed{{}^{10}C_9 \, (0.3)^9 \, (0.7)^1}$

Very small number

↳ Binomial Distribution

↳ pmf → ${}^{n}C_k (P)^k (1-P)^{n-k}$

2. 3 out of 10 students get placed?

$n=10$

$k=3$

$${}^{10}C_3 (0.3)^3 (1-0.3)^{10-3}$$

* If number of trial = 1 then binomial become bernauli

An engineering college has a placement system where any given student has a 0.3 chance of getting placed through campus recruitment, 0.05 chance of opting out of the placement process, and a 0.65 chance of trying but not getting placed. If you randomly select a student, what is the probability that the student:

1. Gets placed? → $\boxed{0.3}$
2. Doesn't get placed but doesn't opt out either? → $\boxed{0.65}$
3. Opts out of placement? → $\boxed{0.05}$

P(placed) = 0.3
P(optout) = 0.05
P(not placed) = 0.65

→ not binary
→ Categorical
↓
more than 2 Categorical

↳ Categorical Distribution / Multinomial Distribution

An engineering college has a placement system where any given student has 0.3 chance of getting placed through campus recruitment, a 0.05 chance of opting out of the placement process, and a 0.65 chance of trying but not getting placed. If you randomly select 10 students what is the probability that: $\boxed{n=10}$

1) 3 students get placed, 1 student opts out of placement, and 6 students try but do not get placed?

2) No ~~student~~ student gets placed, 2 students opt out of placement, and 8 students try but do not get placed?

P(placed) = 0.3

P(opt out) = 0.05

P(not placed) = 0.65

→ 3 placed, 1 opts out, 6 not placed

PPP O NNN NNN

0.3 0.3 0.3 0.5 $(0.65)^6$

$$\rightarrow \frac{n!}{n_1!\, n_2!\, n_3! \ldots n_k!}$$

$$= \frac{10!}{3!\,1!\,6!} (0.3)^3 (0.05)^1 (0.65)^6$$

→ 0 placed, 2 opts out, 8 not placed

$$= \frac{10!}{0!\,2!\,8!} (0.3)^0 (0.05)^2 (0.65)^8$$

[Multinomial Distribution]

→ We multiply this because order may be change.

Order changed { 2opts out 0placed & not pl / & not place 2opt 0placed

* here we want multiple combination

Definitions

Bernoulli Distribution: The Bernoulli Distribution is a discrete probability distribution that models the outcomes of a binary random variable.

Binomial Distribution: The binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of independent Bernoulli trials.

Categorical Distribution: The categorical distribution is a discrete probability distribution that models the probabilities of different outcomes in a categorical or discrete random variable.

Unlike the Bernoulli or binomial distribution that deal with binary outcomes, the categoricals distribution accomodates multiple categories or outcomes. Each category has an associated probability, and the sum of the probabilities for all categories is equal to 1.

Multinomial Distribution: Multinomial Distribution allows us to calculate the probability of observing a specific count or combination of counts for category in a fixed number of trials.

→ 2 red, 6 blue, 2 green type Question

Multinomial

$\boxed{k=2}$, n	k, $n=1$	$k=2$ $n=1$
Binomial	Multinoulli (Categorical)	Bernoulli

<u>Naive Bayes</u>

- Bernoulli NB $\longrightarrow$ Bernoulli
- Categorical NB $\longrightarrow$ Multinoulli
- Multinomial NB $\longrightarrow$ Multinomial

<u>Bernoulli Naive Bayes</u>

Bernoulli NB implement the naive Bayes training and classification algorithms for data that is distributed according to multivariate Bernoulli distribution. i.e. there may be multiple features but each one is assumed to be a binary-valued (Bernoulli, Boolean) variable.

Therefore, this class requires samples to be represented as binary-valued feature vectors; if handed any other kind of data, a bernoulli NB instance may binarize its input (depending on the binarize parameter).

(B)

Data → Data follow bernoulli distribution

f_1	f_2	f_3
1	0	Yes
0	0	No
0	1	No
1	0	Yes

	doc ID	Words in document	in C = Chinese
Training set	1.	chinese Beijing chinese	Yes
	2.	chinese chinese shanghai	Yes
	3.	chinese Macao	Yes
	4.	Tokyo Japan chinese	No
test	5.	chinese chinese chinese Tokyo Japan	?

→ Bag of words ↓ Binary

↓ binary BoW

	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan	output
d_1	1	1	0	0	0	0	Y
d_2	1	0	1	0	0	0	Y
d_3	1	0	0	1	0	0	Y
d_4	1	0	0	0	1	1	N
d_5	1	0	0	0	1	1	?

P(Y/ chinese = 1, bej = 0, shan = 0, Mac = 0, Tok = 1, Jap = 1)

P(N/ chinese = 1, bej = 0, shan = 0, Mac = 0, Tok = 1, Jap = 1)

$P(Y)$ $P(\text{Chinese}=1/Y)$ $P(\text{bey}=0/Y)$ $P(\text{shan}=0/Y)$ $P(\text{mac}=0/Y)$ $P(\text{tok}=1/Y)$ $P(\text{Jap}=1/Y)$

$P(\text{Chinese}=1/Y) = \frac{3}{3} = 1$

$P(X=k) = pk + (1-p)(1-k)$

$k=0,1 \quad p(1) + (1-p).(1-0)$

$= p = 1$

$P(\text{bey}=0/Y) = \quad pk + (1-p)(1-k)$

$\underset{\text{Ysk}}{}$

$P(1)$

$\underbrace{\left(\frac{1}{3}\right)(0)}_{0} + \left(1-\frac{1}{3}\right)(1-0)$

$= \frac{2}{3}$

$p(\text{Shan}=0/Y) = \quad$ also $\frac{2}{3}$

$P(\text{Mac}=0/Y) = \quad$ also $\frac{2}{3}$

$P(\text{Tok}=1/Y) = \frac{0}{3} \rightarrow$ here we use laplace additive smoothing

* If present in Test data then we use $p(1)$
* If it is not present $= (1-p(1))$
* Same Step for 'No'
* laplace additive smoothing in bernoulli $\rightarrow \frac{2+\alpha}{3\ n\alpha}$

$\hookrightarrow$ no. of feature in particular column.

Multinomial Naive Bayes

	Chinese	beijing	Shanghai	Macao	Tokyo	Japan	Output
d_1	2	1	0	0	0	0	Y
d_2	2	0	1	0	0	0	Y
d_3	1	0	0	1	0	0	Y
d_4	1	0	0	0	1	1	N
d_5	3	0	0	0	1	1	?

P(Y / chin=3, Bej=0, shan=0, mac=0, tok=1, Jap=1]

P(Y) P(Chin=3/Y) P(bej=0/Y) P(Shan=0/Y) P(mac=0/Y) P(tok=1/Y) P(Jap=1/Y)

Chinese Chinese Chinese Tokyo Japan

P(Chinese/Y) P(Chinese/Y) P(Chinese/Y) P(Tokyo/Y) P(Japan/Y)

$\frac{5}{8}$ $\frac{5}{8}$ $\frac{5}{8}$ $\frac{0}{8}$ $\frac{0}{8}$

$$\left(\frac{3}{4}\right)\left(\frac{5}{8}\right)^3\left(\frac{0}{8}\right)^1\left(\frac{0}{8}\right)^1$$

prob of Y

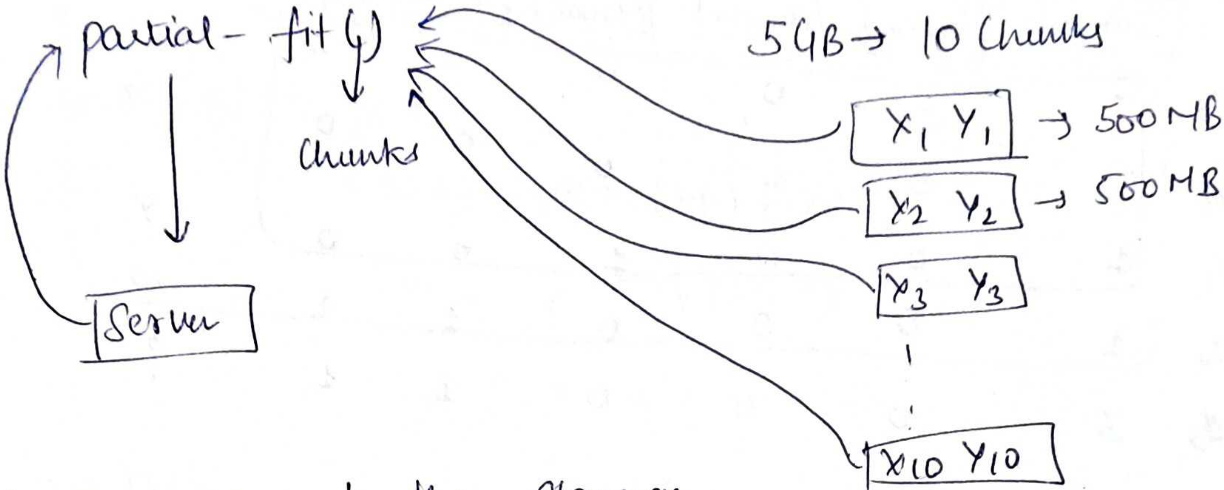
$\frac{5!}{3!\,1!\,1!}$ why not multiply with

* because order is same [Chinese Chinese Chinese Tok Jap]

Can't change order like [Chinese Tok Jap Chinese Chinese]

* we just want One combination.

Out of Core Naive Bayes



* If Data exceed the Storage and RAM.