# Types of NAIVE BAYES in Sklearn

- Gaussion
- Categorical
- Bernouli
- Multinomial
- Complement

# (1) Gaussian Naive Bayes

→ When do use?
- When all features are numerical
- assume all are normal distribution

- Eg: Data ⇒

| cgpa | iq | placed |
|---|---|---|
| – | – | Y |
| – | – | Y |
| – | – | N |
| – | – | N |
| – | – | Y |

(500 – Y, 500 – N)

- Query → { 8.1, 81 } → Y / N
- 2 prob will be calculated
- $\underbrace{P(Y \mid cgpa = 8.1, iq = 81)} = P(Y) \times \underline{P(cgpa = 8.1 \mid Y)} \times \underline{P(iq = 81 \mid Y)}$

- Prob. of underlined terms will be calculated using Gaussian pdf (As per assumption)

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

- $$P(cgpa = 8.1 \mid Y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$ ; $x = 8.1$, $\mu$ = mean of cgpa of Y classes, $\sigma$ = std of cgpa of Y classes

- $$P(iq = 81 \mid Y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{81-\mu}{\sigma}\right)^2}$$, $x = 81$, $\mu$ = mean of iq, $\sigma$ = std of iq

- put all these values you got your $P(Y \mid cgpa = 8.1, iq = 81)$

- Same goes for $P(N \mid gpa = 8.1, iq = 81)$

M T W T F S S
Page No :
Date:
YOUVA

- whosever prob will be highest that label will be assigned to new query

Q Why Laplace Additive Smoothing isn't applied on Gaussian Naive Bayes

Ans. Because, Laplace Additive was applied for preventing probability to become 0.

- But since, we assumed it to be Gaussian and it is a continuous column, for every value there will be probability density

## (2) ~~Multinomial~~ Categorical Naive Bayes

- The basic Naive Bayes, we were studying from the start

- When to use :- when all features are categorical

Eg:

| Outook | Temperature | Playtennis |
| --- | --- | --- |
| Sunny | Hot | No |
| Rainy | cool | Yes |
| ⋮ | | |

- New query → { Sunny, Hot, } ~~High~~ → Yes / No

- $P(\text{Yes} \mid_{\text{Outlook = sunny, Temp = hot}}) = P(\text{Yes}) \times P(\text{sunny} \mid \text{Yes}) \times P(\text{hot} \mid \text{Yes})$

- $P(\text{No} \mid \text{sunny, hot}) = P(\text{No}) \times P(\text{Sunny} \mid \text{No}) \times P(\text{hot} \mid \text{No})$

- In this, you can apply Lapace Additive smoothing

- $P(Yes \mid sunny, heat) = P(Yes) \times \underbrace{P(sunny \mid Y) \times P(heat \mid Y)}$

  LAS will be applied on likelihood, conditional

- same goes for $P(No \mid sunny, heat)$

- let's say this term is $\Rightarrow \frac{3}{5}$

- Applying LAS : $\frac{3 + \alpha}{5 + \alpha n_i} \rightarrow$ n is no of categories in that column

$$= \frac{3 + 1}{5 + 1(3)} = \frac{2}{8}$$

## (3) Multinomial Naive Bayes

**When to use?**

When all features are discrete

Eg Data →

| $f_1$ | $f_2$ | $f_3$ ...... | $f_n$ |
|---|---|---|---|
| 1 | 0 | 2 | 5 |
| 6 | 3 | 1 | 4 |

- Generally Textual data is most common example of Discrete Data
- BOW types → Binary BOW
  → Count BOW
- Works good on Count BOW
- fractions can also be considered (TF-IDF terms)

→ good result on textual data

## Understanding Multinomial with Eg:

data →

| | doc id | words | in c = china |
|---|---|---|---|
| Training | 1 | Chinese Beijing Chinese | Yes |
| | 2 | Chinese chinese shanghai | Yes |
| | 3 | Chinese macao | Yes |
| | 4 | Tokyo japanese Chinese | No |
| Test | 5 | Chinese Chinese chinese tokyo japan | ? |

- **Applying BOW**

→ unique words = [Chinese, Beijing, Shanghai, macao, Tokyo, Japan]

| | 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 |

- We have to predict for { chinese, Chinese chinese tokyo japan
- Transforming it acc to BOW

| | Chinese | Beijing | Shang | Macao | Tokyo | Japan | |
|---|---|---|---|---|---|---|---|
| d5 | 3 | 0 | 0 | 0 | 1 | 1 | ? |

- We have to predict Y or N for d5
- 2 prob will be calculated.

→ $P(Y \mid \text{chinese}=3, \text{Beijing}=0, \text{Shangi}=0, \text{Macao}=0, \text{Tokyo}=1, \text{Japan}=1)$

&

→ $P(N \mid \text{chinese}=3,$ " " " " " " $)$

- To find $P(Yes \mid X)$ & $P(No \mid X)$, we first need to do (calculate) all possible probabilities

Ony Yes

$\rightarrow$ $P(Y) = \frac{3}{4}$, $P(\text{Chinese} \mid Y)$, $P(\text{Bejing} \mid Yes)$ .......... $P(\text{Japan} \mid Y)$

$\rightarrow$ $P(\text{Chinese} \mid Y) = ?$

- It is total no of chinese words in Yes label, divided by total no of words in Yes column

- That means $\Rightarrow \frac{5}{8}$ (Check in data)

$\rightarrow$ $P(\text{Beijing} \mid Y) = \frac{1}{8}$,

$\rightarrow$ $P(\text{Japan} \mid Y) = \frac{0}{8}$ (Similary find for other)

Due to this apply laplace

i.e $\frac{0 + \alpha}{8 + n\alpha}$ — (1)

- What is this $n$?
- $B = n\alpha$
- and $B$ is the size of vocabulary
- $\alpha = 1$ (generally)

$\therefore$ $n = \frac{B \text{ (size of vocab)}}{\alpha}$

$n = \frac{6}{1}$

$\therefore$ $P(\text{Japan} \mid Y) = \frac{0}{8} = \frac{0+\alpha}{8+n\alpha}$

$= \frac{0+1}{8+(6\times1)} = \frac{1}{14}$

- Similary apply for other terms

$P(\text{Chinese} \mid Y) = \frac{5+1}{8+(6\times1)} = \frac{6}{14} = \frac{3}{7}$

Same thing will be applied for NO i.e. find all possibles probabilities before hand

→ $P(\text{Chinese} \mid \text{No}) = \frac{1+1}{3+(6\times1)} = \frac{2}{9}$

→ $P(\text{Tokyo} \mid \text{No}) = \frac{1+1}{3+(6\times1)} = \frac{2}{9}$

→ $P(\text{Japan} \mid \text{No}) = \frac{1+1}{3+(6\times1)} = \frac{2}{9}$

Similarly, other probab for No will be calculated

Now, we had to predict :-

(1) $P(Y \mid \text{Chinese}=3, \text{bei}=0, \text{sha}=0, \text{mac}=0, \text{tok}=1, \text{japan}=1)$

- We find all these probabilitis for yes such as $P(\text{chinese} \mid Y)$, $P(\text{bei} \mid Y)$.

- and our prediction wants chinese = 3, bei = 0
- we will raise the power by these values of our probabilities which we find:

- that means ⇒
  - chinese = 3 ⇒ $P(\text{Chinese} \mid Y)^3$
  - bei = 0 ⇒ $P(\text{bei} \mid Y)^0$
  - sha = 0 ⇒ $P(\text{Sha} \mid Y)^0$
  - mac = 0 ⇒ $P(\text{mac} \mid Y)^0$
  - tok = 1 ⇒ $P(\text{Tok} \mid Y)^1$
  - Jap = 1 ⇒ $P(\text{Jap} \mid Y)^1$

- and all these probabilities we had already found, we just have to raise the powers

- Same goes for $P(N \mid X)$

(2) $P(N \mid \text{chinese}=3, \text{bi}=0, \text{sha}=0, \text{mac}=0, \text{tok}=1, \text{jap}=1)$

$= P(N) \times P(\text{chin} \mid N)^3 \times P(\text{bei} \mid N)^0 \times P(\text{sha} \mid N)^0 \times P(\text{mac} \mid N)^0 \times P(\text{tok} \mid N)^1 \times P(\text{jap} \mid N)^1$

M T W T F S S

| Page No.: | YOUVA |
| --- | --- |
| Date: | |

Q But why we raised it to the power of that query pt values?

A. Due to Multinomial Distribution