Data Science and Artificial Intelligence

Machine Learning

Bayesian learning.

Lecture No. 3

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic — Questions

Topic — Naive Bayes.

Topic

Topic

Topic

Topics to be Covered

Topic

Naive Bayes algorithm

Topic

Topic

Topic

Topic


Stay Strong
BE
Positive

Naïve Bayes Algorithm

Bayes Theorem

Bayes' theorem is the foundation of the Naive Bayes algorithm. It describes the probability of an event based on prior knowledge of conditions that might be related to the event. The theorem is stated as:

Bayes theorem

$$\underbrace{P(A|B)}_{\text{Posterior}} = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Prior

likelihood

Marginal Probab

Where:

- $P(A|B)$ is the posterior probability: the probability of event A occurring given that B is true.
- $P(B|A)$ is the likelihood: the probability of event B occurring given that A is true.
- $P(A)$ is the prior probability: the initial probability of event A.
- $P(B)$ is the marginal probability: the total probability of event B occurring.

Naïve Bayes classifier

In the context of a classifier, Bayes' theorem can be rewritten as:

$$P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)}$$

Where:

- y is the class label.
- X is the feature vector ($X = (x_1, x_2, ..., x_n)$).

Naïve Bayes Algorithm

Naïve Bayes classifier

The "naive" assumption is that all features x_i are conditionally independent given the class label y. This simplifies the computation:

$$P(X|y) = P(x_1|y) \cdot P(x_2|y) \cdot \ldots \cdot P(x_n|y)$$

Thus, the posterior probability becomes:

$$P(y|X) \propto P(y) \cdot \prod_{i=1}^{n} P(x_i|y)$$

The classifier then selects the class with the highest posterior probability.

Bayesian Decision Theory

Naïve Bayes Algorithm

→ Classification

Rule:

$P(C_1/x) \underset{C_1 \quad C_2}{><} P(C_2/x)$ → Posterior Probab Compare

2 Class of for any data Point x

$P_{C_1} . P(x|C_1) \underset{C_1 \quad C_2}{><} P_{C_2} P(x|C_2)$

⇓ assumption

$P_{C_1} P(x^1/C_1) P(x^2/C_1) --- >< P_{C_2} P(x^1/C_2) P(x^2/C_2) ---$

Naïve Bayes Algorithm

The fundamental Naive Bayes assumption is that each feature makes an:

- **Feature independence:** The features of the data are conditionally independent of each other, given the class label.
- **Features are equally important:** All features are assumed to contribute equally to the prediction of the class label.

Naïve Bayes Algorithm

We can convert the MAP into…

Bayesian Decision Theory

Naïve Bayes Classifier

4 dimension | Label

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Test Point:
X = [Sunny, Cool, high, strong]
find label of X.

Label → Categorical → Yes/No
Outlook → " → Sunny, overcast, Rain
Temp → " → Hot, mild, Cool
Humidity → " → high, normal
Wind → " → weak, Strong.

Bayesian Decision Theory

Naïve Bayes Classifier

4 dimension | Label

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Test Point:
X = [Sunny, Cool, high, strong]
find label of X.

{Eq. $P_{C_1} \cdot P(x^1|C_1)P(x^2|C_1) \dots >< P_{C_2} P(x^1|C_2)P(x^2|C_2) \dots$}

$\Rightarrow P_Y \cdot P(\text{Sunny}/Y)\, P(\text{Cool}/Y)\, P(\text{high}/Y)\, P(\text{Strong}/\text{Yes}) >< P_N P(\text{Sunny}/N)\, P(\text{Cool}/N)\, P(\text{high}/N)\, P(\text{strong}/N)$

Bayesian Decision Theory

Naïve Bayes Classifier

4 dimension

Label

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

- In training phase of naive bayes we find all the probabilities using training data

So $P_Y \cdot P(x^1/Y)\,P(x^2/Y)\,P(x^3/Y)\,P(x^4/Y)$

$P_N \cdot P(x^1/N)\,P(x^2/N)\,P(x^3/N)\,P(x^4/N)$

Sunny
Overcast
Rain

Hot
mild
Cool

High
Normal

weak
Strong.

Bayesian Decision Theory

Naïve Bayes Classifier

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

No of times we get Sunny / Total No of points

Outlook	P(O/Yes)	P(O/No)
Sunny	$P(S/Y) = 2/9$	
Overcast	$P(O/Y) = 4/9$	
Rain	$P(R/Y) = 3/9$	

- So here label is given Yes
- So Consider points with Yes label

Bayesian Decision Theory

a Yes b No

Naïve Bayes Classifier

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

No of times we get Sunny / Total No of points

Outlook	P(O/Yes)	P(O/No)
Sunny	$P(S/Y) = 2/9$	$P(S/N) = 3/5$
Overcast	$P(O/Y) = 4/9$	$P(O/N) = 0$
Rain	$P(R/Y) = 3/9$	$P(R/N) = 2/5$

- So here label is given Yes
- So Consider points with Yes label

Consider No Points

Bayesian Decision Theory

Naïve Bayes Classifier

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Temperature	P(T/Yes)	P(T/No)
Hot	$P(H/Y) = 2/9$	$P(H/N) = 2/5$
Mild	$P(M/Y) = 4/9$	$P(M/N) = 2/5$
Cold	$P(C/Y) = 3/9$	$P(C/N) = 1/5$

Bayesian Decision Theory

Naïve Bayes Classifier

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Humidity	P(H/Yes)	P(H/No)
High	$P(H/Y) = 3/9$	$P(H/N) = 4/5$
Normal	$P(N/Y) = 6/9$	$P(N/N) = 1/5$

Bayesian Decision Theory

Naïve Bayes Classifier

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak ✓	Yes
Rain	Mild	High	Weak ✓	Yes
Rain	Cool	Normal	Weak ✓	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak ✓	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak ✓	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak ✓	Yes
Rain	Mild	High	Strong	No

Wind	P(W/Yes)	P(W/No)
Weak	$P(\omega/Y) = 6/9$	$P(\omega/N) = 2/5$
Strong	$P(S/Y) = 3/9$	$P(S/N) = 3/5$

- $P_Y \Rightarrow 9/14$
- $P_N \Rightarrow 5/14$

directly observing data

Bayesian Decision Theory

Training of Naive bayes

$P_Y = 9/14$
$P_N = 5/14$

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Outlook	P(O/Yes)	P(O/No)
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	P(T/Yes)	P(T/No)
Hot	2/9	2/5
Mild	4/9	2/5
Cold	3/9	1/5

Humidity	P(H/Yes)	P(H/No)
High	3/9	4/5
Normal	6/9	1/5

Wind	P(W/Yes)	P(W/No)
Weak	6/9	2/5
Strong	3/9	3/5

Bayesian Decision Theory

$P_Y = 9/14$
$P_N = 5/14$

Outlook	P(O/Yes)	P(O/No)
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	P(T/Yes)	P(T/No)
Hot	2/9	2/5
Mild	4/9	2/5
Cold	3/9	1/5

Humidity	P(H/Yes)	P(H/No)
High	3/9	4/5
Normal	6/9	1/5

Wind	P(W/Yes)	P(W/No)
Weak	6/9	2/5
Strong	3/9	3/5

- After training we Remove data and only store these probabilities

Test point :- X [Sunny, mild, high, weak] → No

$\rightarrow P_Y P(S|Y) P(m|Y) P(h|Y) P(\omega|Y) = \frac{9}{14} \cdot \frac{2}{9} \cdot \frac{4}{9} \cdot \frac{3}{9} \cdot \frac{6}{9}$

$= .01410$

$P_N P(S|N) P(m|N) P(h|N) P(\omega|N) = \frac{5}{14} \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{2}{5}$

$= .0274$

Bayesian Decision Theory

$P_Y = 9/14$
$P_N = 5/14$

- After training we Remove data and only store these probabilities

Test point :- X [Rain, Cold, $\boxed{\text{No}}$ High, Strong]

$P_Y P(R|y) P(C|y) P(H|y) P(S|y) \Rightarrow \frac{9}{14} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9}$

$\Rightarrow 7.93 \times 10^{-3}$

$P_N P(R/N) P(C/N) P(H/N) P(S/N) \Rightarrow \frac{5}{14} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5}$

$\Rightarrow .013714$

Outlook	P(O/Yes)	P(O/No)
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	P(T/Yes)	P(T/No)
Hot	2/9	2/5
Mild	4/9	2/5
Cold	3/9	1/5

Humidity	P(H/Yes)	P(H/No)
High	3/9	4/5
Normal	6/9	1/5

Wind	P(W/Yes)	P(W/No)
Weak	6/9	2/5
Strong	3/9	3/5

Bayesian Decision Theory

$P_Y = 9/14$
$P_N = 5/14$

Outlook	P(O/Yes)	P(O/No)
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	P(T/Yes)	P(T/No)
Hot	2/9	2/5
Mild	4/9	2/5
Cold	3/9	1/5

Humidity	P(H/Yes)	P(H/No)
High	3/9	4/5
Normal	6/9	1/5

Wind	P(W/Yes)	P(W/No)
Weak	6/9	2/5
Strong	3/9	3/5

- After training we Remove data and only store these probabilities

Test point :- X [Overcast, mild, high, weak] → Yes.

→ $P_Y P(O/Y) P(m/y) P(h/y) P(w/y)$ → Non zero

→ $P_N \underbrace{P(O/N)}_{0} P(m/N) P(h/N) P(w/N)$ → 0

Bayesian Decision Theory

$P_Y = 9/14$
$P_N = 5/14$

Outlook	P(O/Yes)	P(O/No)
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	P(T/Yes)	P(T/No)
Hot	2/9	2/5
Mild	4/9	2/5
Cold	3/9	1/5

Humidity	P(H/Yes)	P(H/No)
High	3/9	4/5
Normal	6/9	1/5

Wind	P(W/Yes)	P(W/No)
Weak	6/9	2/5
Strong	3/9	3/5

- After training we Remove data and only store these probabilities

Test point :- X [Overcast, ✓, ✓, ✓]

→ always Y class.

→ (Zero probability problem)

Naïve Bayes Algorithm

- Problems in Naive Bayes Algo ⇒ → need large dataset.

1. Zero probability problem ⇒ lack of data,
2. Here we are Calculating probabilities of a dimension given a class, to get exact value of " we need large dataset

Naïve Bayes Algorithm

Complexity in naïve bayes classfier…

Naïve Bayes Algorithm

Consider the following dataset showing the result whether a person has passed or failed the exam based on various factors. Suppose the factors are independent to each other. We want to classify a new instance with Confident=Yes, Studied=Yes, and

Test : Yes Yes No

Find Result
a) P
b) F

$P_{Pass} = 3/5,\ P_{Fail} = 2/5$

$\to P_P\, P(Y/P)\, P(Y/P)\, P(N/P) \Rightarrow \frac{3}{5} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \Rightarrow .088$

$\to P_F\, P(Y/F)\, P(Y/F)\, P(N/F) = \frac{2}{5} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \Rightarrow .05$

Confident	Studied	Sick	Result
Y	N	N	Fail
Y	N	Y	Pass
N	Y	Y	Fail
N	Y	N	Pass
Y	Y	Y	Pass

Naïve Bayes Algorithm

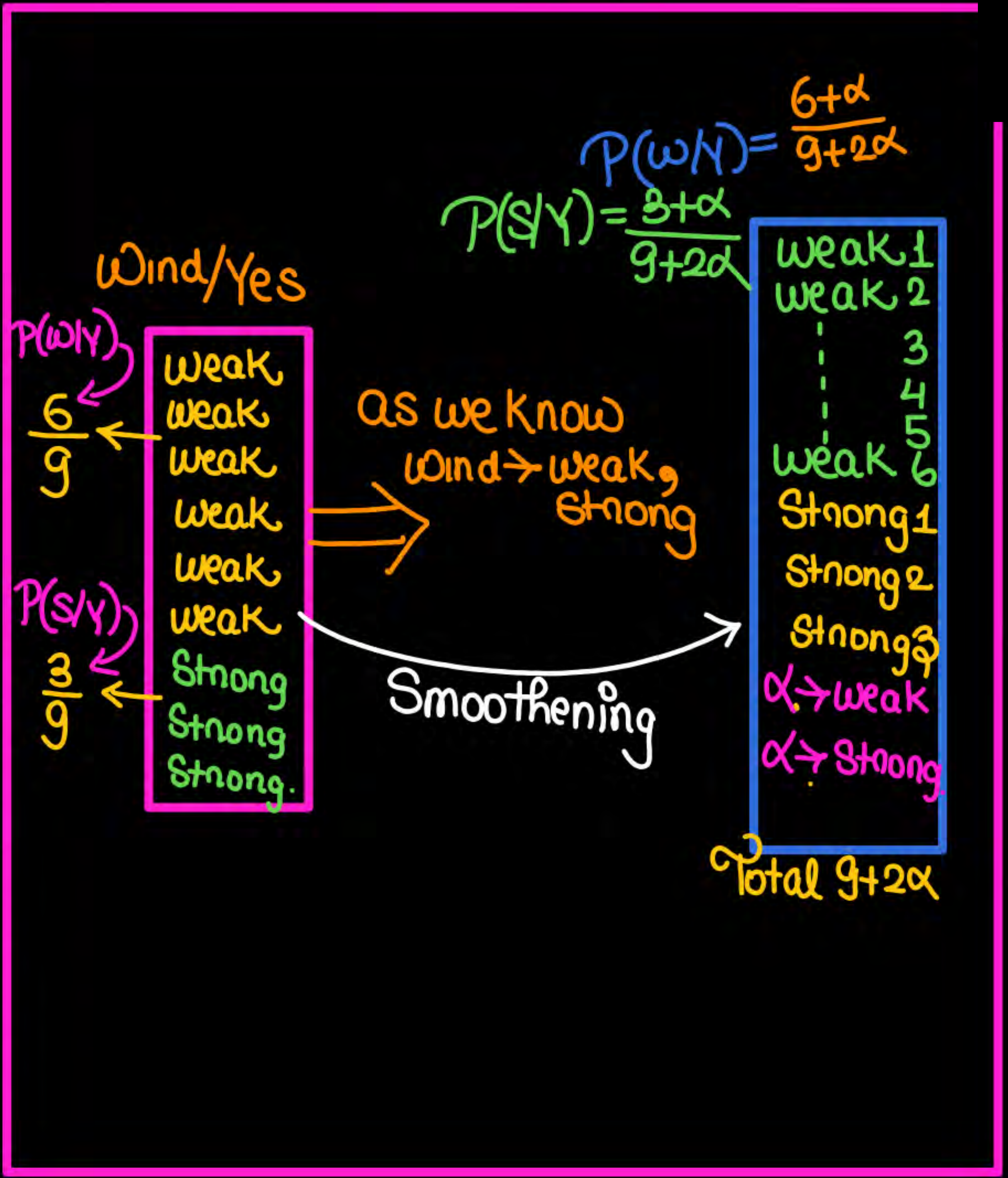
Zero probability problem…

P(value of a dimension | class) = 0

generally due to lack of data

Bayesian Decision Theory

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No



Bayesian Decision Theory

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

$$P(\omega/N) = \frac{2+\alpha}{5+2\alpha}$$

$$P(S/N) = \frac{3+\alpha}{5+2\alpha}$$

Wind/NO

$P(\omega/N) = 2/5$

$P(S/N) = 3/5$

weak
weak
Strong
Strong
Strong

as we know
wind → weak, strong

Smoothening

weak 1
weak 2
Strong 1
Strong 2
Strong 3
α → weak
α → strong

Total = $5+2\alpha$

We find α by CrossValidation

Bayesian Decision Theory

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak ✓	No
Sunny	Hot	High	Strong	No
Overcast ✓	Hot	High	Weak	Yes
Rain ✓	Mild	High	Weak	Yes
Rain ✓	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast ✓	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak ✓	No
Sunny ✓	Cool	Normal	Weak	Yes
Rain ✓	Mild	Normal	Strong	Yes
Sunny ✓	Mild	Normal	Strong	Yes
Overcast ✓	Mild	High	Strong	Yes
Overcast ✓	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Outlook/Yes

Sunny 2
OverCast 4
Rain 3

$P(S/Y) = 2/9$
$P(O/Y) = 4/9$
$P(R/Y) = 3/9$

Outlook
S O R

$P(S/Y) = \frac{2+\alpha}{9+3\alpha}$
$P(O/Y) = \frac{4+\alpha}{9+3\alpha}$
$P(R/Y) = \frac{3+\alpha}{9+3\alpha}$

Sunny 2
OverCast 4
Rain 3
Sunny α
OverCast α
Rain α

Bayesian Decision Theory

Outlook	Temp.	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Weak	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Strong	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Outlook/No

Sunny 3
Overcast 0
Rain 2

$P(S/N) = 3/5$

$P(O/N) = 0$

$P(R/N) = 2/5$

Outlook

S O R

$P(S/N) = \frac{3+\alpha}{5+3\alpha}$

$P(O/N) = \frac{0+\alpha}{5+3\alpha}$

$P(R/N) = \frac{2+\alpha}{5+3\alpha}$

Sunny 3
Overcast 0
Rain 2
Sunny α
Overcast α
Rain α

Naïve Bayes Algorithm

How to Smoothen data ⇒

↳ we donot inc the number of datapoints

→ The smothening is done dimension wise

→ dimension/class

→ wind/yes

Solving the zero-probability problem…

→ (i) increase data

(ii) Smoothing technique

THANK - YOU