NLP

Deep learning

well

'NN'

Project

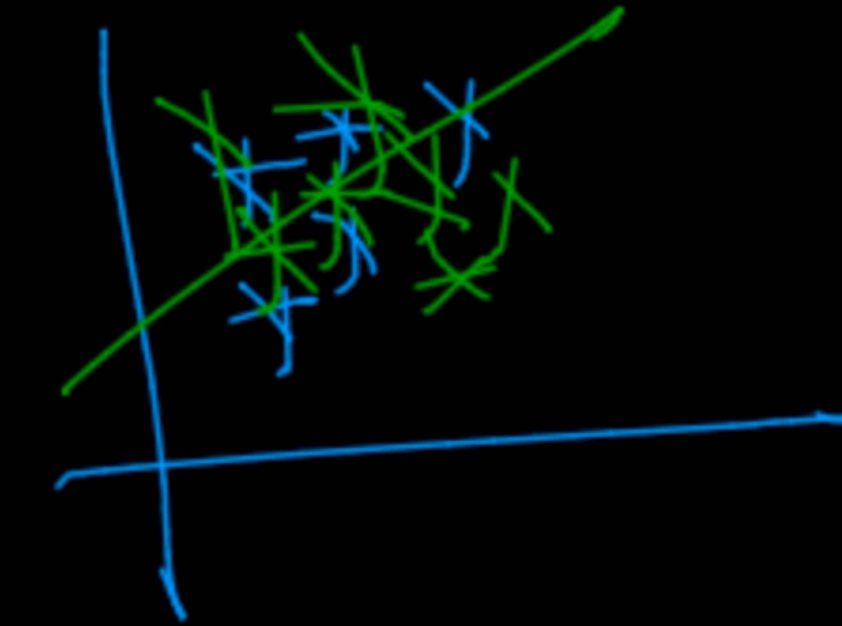
sum

text → vector

Naive Bayes

Algorithm

email
↓
vectorised



logR

Y

X

spam ← Email → ML → ML'

collect

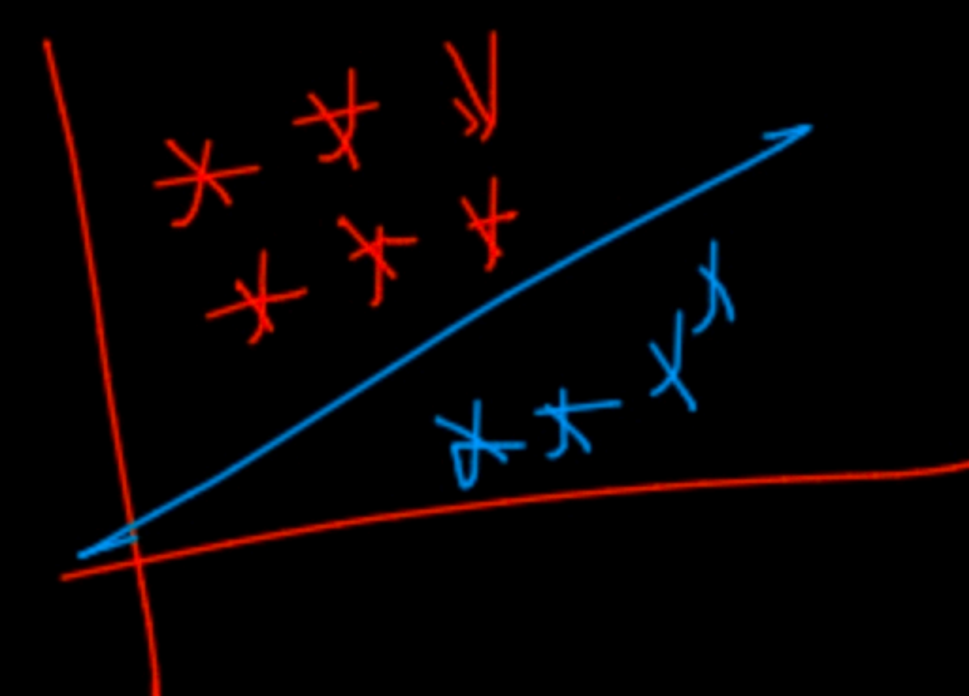
lin

nm

10K

'Log Regression'

spam / NOT



(Acc ↓)

text

Emain → vector 'tfidf'

nm

Reviews → □ → (+ve / -ve)

Youtube

topic → collects

comments

vectorisation → tf idf

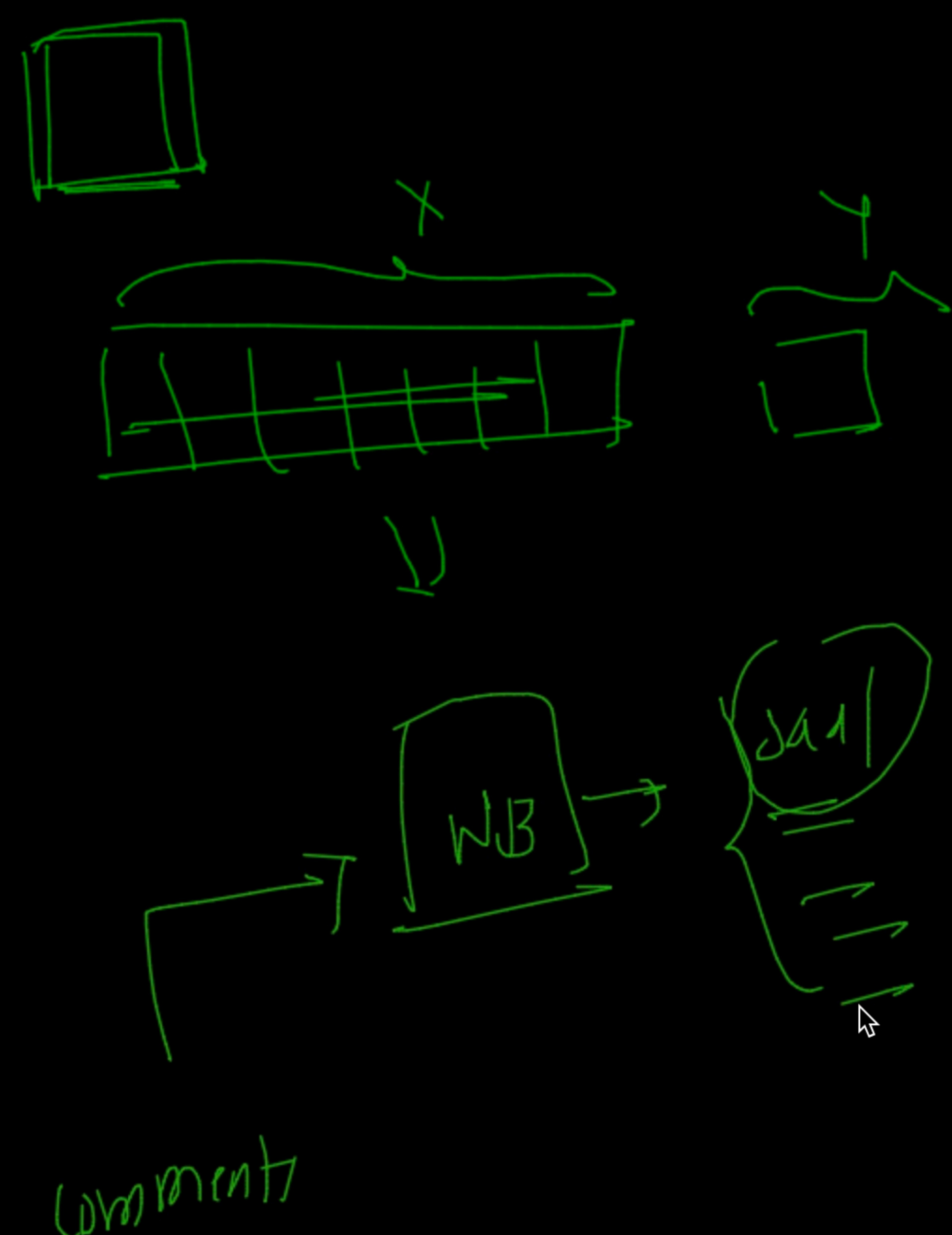
(3-4)

label

sad

An

lot



2 / 2

Probability

⇒ conditional prob (wiki)

A

B

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} \qquad P(B) \neq 0$$

already

occured

$P(B)$

Bayes theorem

$\Downarrow$

likelihood

prior

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

evidence

posterior

What is prob of getting A if B already happend

1 = spam
0 = Not spam

Data	Class
x_i text	1
x_j	0

$$P(y=1 \mid text_q) = ?$$

$$P(y=0 \mid text_q)$$

text → vector data

→ $\{w_1, w_2, w_3 - \quad - - w_d\}$

$$P(Y=1 \mid w_1 w_2 w_3 - - - - w_d)$$

$$= P(Y=1) * P(w_1 \mid Y=1) * P(w_2 \mid Y=1) * - - - P(w_d \mid Y=1)$$

$$= \boxed{P(Y=1) * \prod_{i=1}^{d} P(w_i \mid Y=1)}$$

Very similar

$$P(Y_0 \mid \text{tenda}) = P(Y_0) \prod_{i=1}^{n} P(w_i \mid Y=0)$$

$$P(Y=0) = \frac{\#\ \text{train pts with } Y=1}{\text{total}\ \#\ \text{train pts}}$$

eg

0 → Not spam

1 → spam

training

Data	label
w1 w2 w2	0
w1 w2 w1 w5	0
w3 w4 w1	0
w5 w1 w1	1
w1 w2 w3, w4	1

at time of training

$$P(Y=0) = \frac{3}{5} \Rightarrow 0.6$$

$$P(Y=1) = \frac{2}{5} \Rightarrow 0.4$$

Unique

$$P(w1 \mid Y=1) = \frac{\#\text{ data pts which contain } w1 \text{ \& } y=1}{\#\text{ data points with } y=1}$$

	Y=1	Y=0
w1	1	~~2/3~~ 0.6
w2	1/2	2/3
w3	1/2	1/3
w4	1/2	1/3
w5	1/2	1/3

Bayes

$$P(Y=1 \mid w_1 w_2 w_3)$$

$$= P(Y=1) * P(w1 \mid Y=1) * P(w2 \mid Y=1) * P(w3 \mid Y=1) * w6$$

testing → text → vector → | w1 | w2 | w3 | w6 |

$$0.4 * 1 * 1/2 * 1/2$$

at time of 'training

$P(Y=1)$

$P(w_1|Y=1)$

⋮

$P(w_d|Y=1)$

$P(Y=0)$

$P(w_1|Y=0)$

⋮

$P(w_d|Y=0)$

→ at time of testing

⇓

text$_q$ ⇒

$\{w_1\ w_2\ w_3,\ w'\}$

'new word'

10 / 10

$$P(y=1 \mid \text{text}) = P(y=1) \times P(w_1 \mid y=1) * P(P \mid w_2 \mid y=1)$$

$$* P(w_3 \mid y=1) * P(w' \mid y=1) \rightarrow 1$$

'10'

$$P(y=0 \mid \text{text}) =$$

'10'

$$P(w' \mid y=1)$$

$$P(w' \mid y=1) \Rightarrow \frac{\#\text{time } w' \text{ occur in train \& } y=1}{\#\text{ pts when } y=1}$$

0

0.0 - - - 1

0.14 - - -

$$\frac{0}{n} \rightarrow 0$$

Laplace smoothing

$$P(w' \mid Y=1) = \frac{0+\alpha}{n_1 + K\alpha}$$

$K \Rightarrow$ #distinct value w' can take

$$\boxed{K=2}$$

(true) → 1

$n_1 = 100$

(case ii)

$(\alpha \uparrow)$

1 → 1/2

$$P(\omega^1 \mid Y=1) = \frac{0+\alpha}{100+2\alpha}$$

$$= \frac{\alpha}{100+2\alpha}$$

$\alpha = 10{,}000$

$$\Rightarrow \frac{0 + 10{,}000}{100 + 20{,}000}$$

50%

$$\Downarrow$$

$$\frac{10000}{20000} = \frac{1}{2}$$

(good)

(case i) $\alpha = 1 \rightarrow \dfrac{1}{100+2} \not\Rightarrow 0$

13 / 14

$w_1 \rightarrow -ve$

w_1 — $-ve$

$w_2 \rightarrow +ve$

w_L

$w_1 \rightarrow 50\% \rightarrow (+w)$

small

α

$\uparrow$

Hyper-parameters

$\alpha \uparrow$

$\neq (0)$

$$P(w_i \mid y=1) = \frac{\#\text{data points with } w_i \ \& \ y=1 + \alpha}{\#\text{data points } y=1 + \underline{\alpha} K}$$

α

Naive Bayes $\longrightarrow$ Sklearn