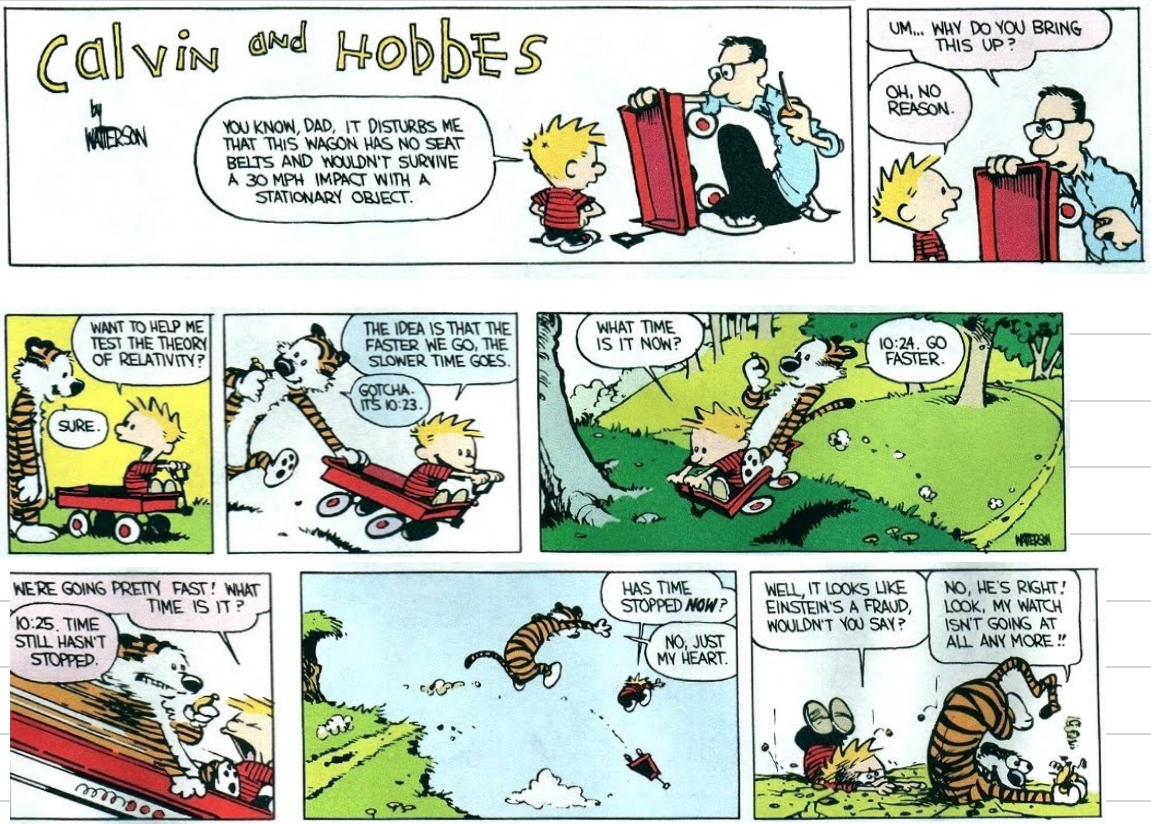
Session - 9

Naive Bayes - 1

Sep 12, 2025



Agenda

1. Naive Bayes
2. Text Processing
3. Assumption
4. Mathematical Intuition

Spam Classifier

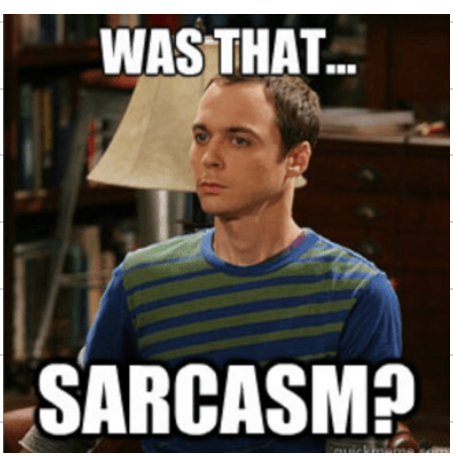
We have found $8 million in transit from India to Russia with your name on the package.

Word-list: ["Nigerian Prince", "Lotter", "Prize", "won",]

$O(n^2)$

Why people love Naive Bayes?

1. Interpretable.
2. Simple
3. Fast
4. Works quite well, for small use-cases.



tmt	Y
	0
	1
	0
	0
	1

Prob of tmt being spam or ham

$P(Y=1 / tmt)$

$P(Y=0 / tmt)$

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

tmt = "You (w_1) won (w_2) Lottery (w_3)" ; N.B. does not care about Sequence

$P(Y=1 / \text{"You won Lottery"})$

&

$P(Y=0 / \text{"You won Lottery"})$

$$P(y=1 / w_1 \cap w_2 \cap w_3) = \frac{P(w_1 \cap w_2 \cap w_3 / y=1) * P(y=1)}{P(w_1 \cap w_2 \cap w_3)}$$

$$P(y=0 / w_1 \cap w_2 \cap w_3) = \frac{P(w_1 \cap w_2 \cap w_3 / y=0) * P(y=0)}{P(w_1 \cap w_2 \cap w_3)}$$

Eq1 $\dfrac{(x^2+y^2+z^{2/2}-\phi)^{3/4}}{(a-b)^2}$

Eq7 $\dfrac{(x^2+z^2+p^2+n^2+r^{3/2}+\phi^{1/2})^{3/4}}{(a-b)^2}$

← tents →			y_i
ω_1	ω_2	ω_3	1
ω_2	ω_3	ω_4	1
ω_5	ω_6	ω_7	1
ω_1	ω_2	ω_1	1
ω_{11}	ω_{12}	ω_{13}	0
ω_{12}	ω_{11}	ω_{15}	0

$P(\omega_1 \cap \omega_2) = 2/6$

$P(\omega_1 \cup \omega_2) - 3/6$

$P(y=1) = 4/6 \quad P(y=0) = 2/6$

$P(\omega_1 / y=1) = 2/4$

$P(\omega_{11} / y=0) = 2/2 = 1$

Let's say:
the probability of dangerous fires are rare (1%)
but smoke is fairly common (10%) due to barbecues,
and 90% of dangerous fires make smoke
Can you find the probability of dangerous Fire when there is Smoke?

0 users have participated

✓	A	0.09	0%
	B	0.9	0%
	C	0.1	0%

End Quiz Now

$= 0.09$

$P(F) = 0.01$ $\quad P(S|F) = 0.9$

$P(S) = 0.1$

$$P(F|S) = \frac{P(S|F) \times P(F)}{P(S)}$$

The Naive Assumption

$$P(w_1 \cap w_2 \cap w_3 / y=1)$$

$$P(A \cap B) = P(A) \times P(B)$$

Naive Assumption: It assumes all words are "conditionally" independent.

$$P(w_1 \cap w_2 \cap w_3 / y=1) = P(w_1 / y=1) \times P(w_2 / y=1) \times P(w_3 / y=1)$$

~~$P(w_1 \cap w_2 \cap w_3) = P(w_1) \times P(w_2) \times P(w_3)$~~

$$P(y=1 \mid w_1 \cap w_2 \cap w_3) = P(w_1 \mid y=1) \times P(w_2 \mid y=1) \times P(w_3 \mid y=1) \times P(y=1)$$

$$P(y=1 \mid w_1 \cap w_2 \cap w_3) = P(w_1 \cap w_2 \cap w_3 \mid y=1) \times P(y=1)$$

$$\prod_{i=1}^{n} P(w_i \mid y=1) \times P(y=1)$$

$$\sum_{i=1}^{3} n_i = n_1 + n_2 + n_3$$

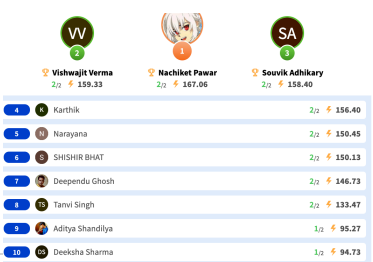
$$\prod_{i=1}^{3} n_1 \times n_2 \times n_3$$

Naive assumption is 'In a sentence, given the class label, words are independent of each other':

0 users have participated

A	False	0%
B ✓	True	0%

End Quiz Now



Which of the following statement is TRUE about the Bayes classifier?

0 users have participated

A	Bayes classifier is an unsupervised learning algorithm.	0%
B ✓	Bayes classifier works on the Bayes theorem of probability.	0%
C	Bayes classifier is a Reinforcement Learning Algorithm	0%
D	It assumes the independence between the independent variables or features.	0%

End Quiz Now

You won lottery : spam
Call me asap: ham

~~$P(\text{You}) \times P(\text{won}) \times P(\text{lotte})$~~

$P(\text{You} \mid y=1) \times P(\text{won} \mid y=1) \times P(\text{lotte} \mid y=1)$

Train and Test Time complexity - Naive Bayes — $N \cdot W$

Train: $\forall_{w_i}$ $P(w_i \mid y=1)$ & $P(w_i \mid y=0)$

spam_dict = {}
ham_dict = {}

for row in dataset:
  for word in row:
    Update dicts

spam_dict['hello'] += 1

LAPLACIAN SMOOTHING

$$P(w_i \mid y=1) = \frac{P(w_i \cap y=1) + \alpha}{P(y=1) + C \times \alpha}$$

C → # of Class

$\alpha \rightarrow$ Overfitting / Underfitting

In a naive Bayes algorithm, when an attribute value in the testing record has no example in the training set, then the entire posterior probability will be zero.

0 users have participated

A	True	0%
B	False	0%

End Quiz Now