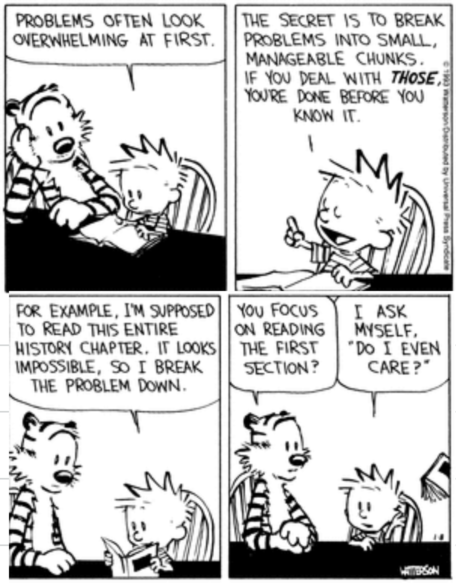
Session - 10

Naive Bayes - 2

Sep 15, 2025



Agenda

1. Training & Test time & Space complexity
2. Hyper-parameter tuning
3. Underflow problem
4. Feature Importance
5. Multinomial Bayes theorem.

Train and Test Time complexity - Naive Bayes — H.W

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

spam_dict = {}
ham_dict = {}

For row in dataset : N
For word in row : D
Update dicts

spam_dict ['hello'] += 1

$O(N \times D \times 2)$

You won $$ 5 trillion = Spam

$P(\text{you} / y=1) = \dfrac{\text{spam\_dict [you]}}{\text{len(rows) where } y=1}$

Test: We've K words in test.

" I love weekend "

For word in test_row: Iterate K time

spam_score = spam_score x spam_dict [word] / (y=1)
ham_sco. = ham_score x ham_dict [word] (y=0)

Test time comp. $O(k)$

I Love Life

$$\text{Spam} = \frac{\text{spam\_dict}[I]}{\text{len}(y=1)} \times \frac{\text{spam\_dict}[\text{Love}]}{\text{len}(y=1)} \times \frac{\text{spam\_dict}[\text{Life}]}{\text{len}(y=1)} \times P(y=1)$$

$$\text{ham} = \frac{\text{notspam\_dict}[I]}{\text{len}(y=0)} \times \frac{\text{notspam\_dict}[\text{Love}]}{\text{len}(y=0)} \times \frac{\text{notspam\_dict}[\text{Life}]}{\text{len}(y=0)} \times P(y=0)$$

Laplace Smoothing

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

$C \to$ # of class

α – hyperparamet

$$P(\text{book} \cap y=1) = 50$$

$$P(\text{mulchand} \cap y=1) = 0$$

Case 1: $\alpha = 1$ million

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

$C \to$ # of class

book: $\dfrac{50 + \alpha}{P(y=1) + 2 \times \alpha} = \dfrac{1 \text{ mil} + \phi}{2 \times 1 \text{ mil} \quad \phi} = \dfrac{1}{2} = 0.5$

mulchand $= \dfrac{0 + \alpha}{P(y=1) + 2 \times \alpha} \approx \dfrac{1}{2} = 0.5$

high bias $\leftarrow$ $\alpha \to$ underfitt

Case 1: $\alpha = 10^{-2}$

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

C → # of class

book: $\dfrac{50 + \alpha}{P(y=1) + 2\times\alpha} = \dfrac{50 + 0.01}{P(y=1) + 2\times 0.01}$

$\simeq \dfrac{50}{P(y=1)}$

mulchand $= \dfrac{0 + 0.01}{P(y=1) + 2\times 0.01} \longrightarrow 0$

high variance.

Impact of Imbalance on Naive Bayes

$P(y=1) = 0.3$; spam

$P(y=0) = 0.7$; not-spam

$$P(y=1 / \text{test}) = \underbrace{P(y=1)}_{0.3} \times \underbrace{\prod_{i=1}^{K} P(w_i / y=1)}_{0.6} = 0.18$$

$$P(y=0 / \text{test}) = \underbrace{P(y=0)}_{0.7} \times \underbrace{\prod_{i=1}^{K} P(w_i / y=0)}_{0.3} = 0.21$$

Some Solution:

1. Dangerous Solution and very context-dependent; remove priors.
2. Over-sampling, If the ratio is very stark, like 0.05 : 0.995, I can try to make it less stark, like 1:3

Will Over-sampling, just affect prior? Or also affect likelihood?

$w_i \cap y_i = 50$

Class minority = 500

" majority = 1000

Before rebalancing

$$P(w_j | y=1) = \frac{50 + \alpha}{500 + 2\times\alpha}$$

Post Rebalancing.

$$\frac{(50 + 50) + \alpha}{(500 + 500) + 2\times\alpha}$$

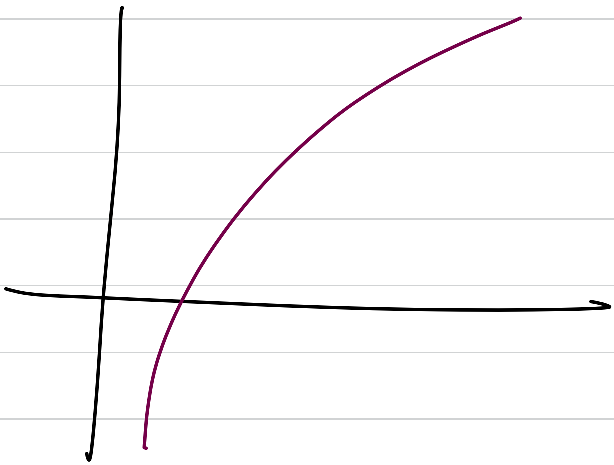
UnderFlow

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

$$0.01 \times 0.01 = 0.0001$$

$$\log(a \times b) = \log(a) + \log(b)$$

$$\log\left(P(y=1) \times \prod_{i=1}^{n} P(w_i \mid y=1)\right) = \log(P(y=1)) + \log(P(w_1 \mid y=1)) + \log(P(w_2 \mid y=1))$$



$$\text{spam} = \log\left(P(y=1) \times \prod_{i=1}^{n} P(w_i \mid y=1)\right)$$

$$\text{ham} = \log\left(P(y=0) \times \prod_{i=1}^{n} P(w_i \mid y=0)\right)$$

Feature Importance and Interpretability

Spam-dict $\forall w_i$ - $P(w_i / y=1)$

ham-dict $\forall w_i$ - $P(w_i / y=0)$

I want to go boating in canals of venice.

While training in Naive bayes, I typically include only top 5k words as part of dataset, so outlier in general is not a concern.

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

$P(w_1 / y=1) = 2/4$

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

:ok:

← texts →			y_i
w_1	w_2	w_3	1
w_2	w_3	w_4	1
w_5	w_3	w_3	1
w_1	w_2	w_1	1
w_{11}	w_{12}	w_{13}	0
w_{12}	w_{11}	w_{15}	0

Multinomial N.B

$P(w_1 / y=1) = 3/4$

$P(w_3 / y=1) = 4/4 = 1$

$$\frac{P(w_i \cap y)}{\sum \text{words } y=1 + V}$$

	w_1	w_2	w_3	- - —	w_n
Sent1	2	1	3		0
Sent2	0	2	1		1

Variance: Error the model makes because of making very complicated assumption/decision boundary.

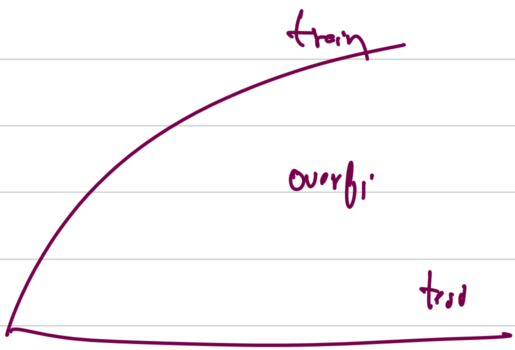
Bias: Error model makes because of making very simple assumption/decision boundary.

$\text{Fish} = 250\,\text{gm} \rightarrow$ Rohu

$250.01\,\text{gm} \rightarrow$ Salmon

$249\,\text{gm} \rightarrow$ Tuna

Fish — Rohu



$\alpha = 1$ million