Naive Bayes

1. Intro
2. Spam data
3. Mathmetical Intution
4. Naive Bayes Assumption
5. Train / Test time complexities
6. Space complexity

Google

Mail 1 : " I, Nigerian prince need your help. Send money "

Mail 2 : " Meeting scheduled at 6 pm. Kindly revert "

lottery , Million dollar , Jackpot

Sentiments

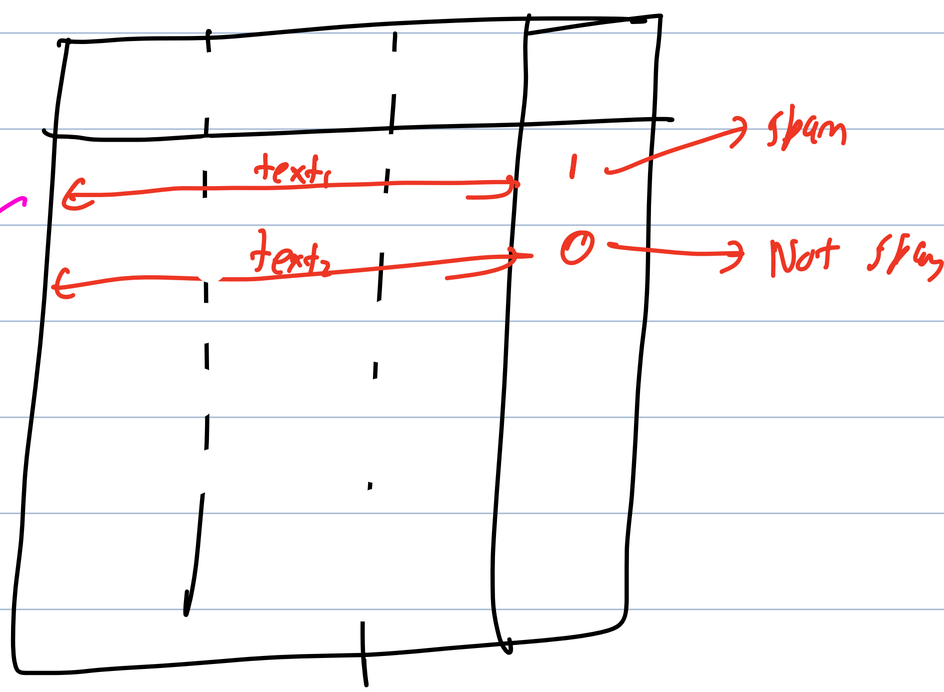
"terrible" , "bad" , "pathetic" → -ve

"great" , "good" → +ve

NB
80%

SOTA
↓
transformers (later)
98.5%

⇒ "I am shivank"
w_1 w_2 w_3



$text_1 = w_1^i \; w_2^i \; w_3^i \; .... \; w_d^i$

"This is shivank"

↓

① Tokenisation ["This", "is", "shivank"]

good Good

② Lowercase

⇒ I, Shivank Agrawal wants to teach!

⇓

removing all punctuation ⇒ regex

④ text → remove stopwords

"I", "and", "the"

⑤ grametical errors ←

Lema Stemming ⇒ eating eat

money

play plays playing

$p(\text{money} \mid \underline{\text{spam}})$

$p(\text{monex} \mid \underline{\text{ham}})$

I eat food

I eaten food

I eats foods

I eat food

Mathematical Intution

$\text{email}_1 \longrightarrow$ spam or Not spam

$\text{email}_1 \longrightarrow 0/1$

$$P(y=1 \mid \text{email}_1)$$

$$P(y=0 \mid \text{email}_1)$$

$$P(y=0 \mid w_1^i, w_2^i, w_3^i, w_4^i, \ldots, w_d^j)$$

$$P(y=1 \mid w_1^i, w_2^i, w_3^i, w_4^i, \ldots, w_d^i)$$

Bayes Theorem

$$\boxed{P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}}$$

$$P(y=1 \mid w_1^i, w_2^i, w_3^i, w_4^i, \ldots, w_d^i)$$

$$P_1(y=1 \mid x) = \frac{P(x \mid y=1) * P(y=1)}{P(x)}$$

$$P_2(y=0 \mid x) = \frac{P(x \mid y=0) * P(y=0)}{P(x)}$$

40/3 ✓

20/3 ✓

$\boxed{y=1}$

$$P(y=1 \mid w_1 \dots w_d) = \frac{P(w_1 \dots w_d \mid y=1) \cdot P(y=1)}{\cancel{K}}$$

$$P(y=0 \mid w_1 \dots w_d) = \frac{P(w_1 \dots w_d \mid y=0) \cdot P(y=0)}{\cancel{K}}$$

$$P(DF) = 0.01$$

$$P(smoke) = 0.1$$

$$P(smoke \mid DF) = 0.9$$

$$P(DF \mid smoke) \Rightarrow ? \Rightarrow \frac{P(DF) * P(smoke \mid DF)}{P(smoke)}$$

$$\Rightarrow \frac{0.01 * 0.9}{0.1}$$

$$\Rightarrow 0.09$$

$$P(y=1 \mid w_1 \dots w_d) = \frac{P(w_1, \dots w_d \mid y=1) \cdot P(y=1)}{\cancel{K}}$$

$$P(y=0 \mid w_1 \dots w_d) = \frac{P(w_1, \dots w_d \mid y=0) \cdot P(y=0)}{\cancel{K}}$$

$$P(y=1) \rightarrow \frac{\#\text{ train points with } y_i = 1}{\text{total }\#\text{ of train pts}}$$

$$P(y=0) \rightarrow \frac{\#\text{ of train point with } y=0}{\text{total }\#\text{ of train pts}}$$

prince | spam

viagra | spam

$$P(\underset{\text{prince}}{w_1}, \underset{\text{viagra}}{w_2} \mid \text{spam}) \approx P(w_1 \mid \text{spam}) \cdot P(w_2 \mid \text{spam})$$

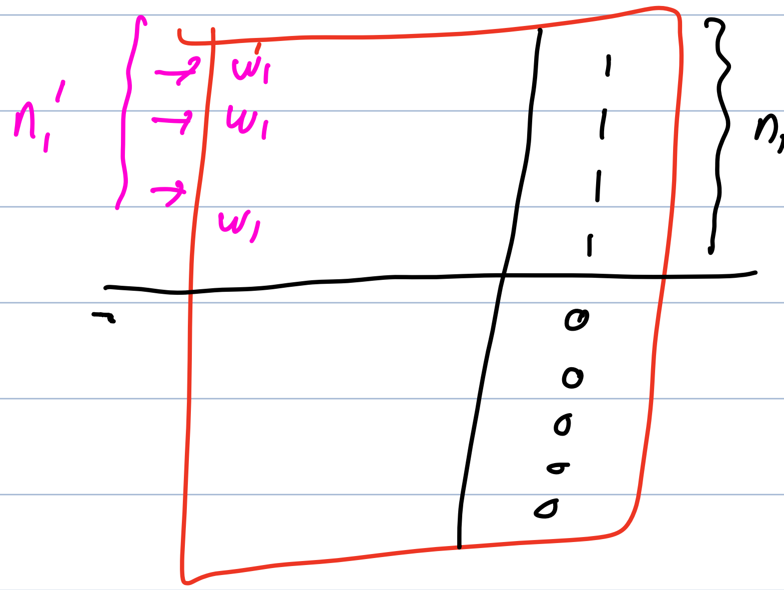
$$P(w_1, w_2 | y=1) = p(w_1 | y=1) * p(w_2 | y=1)$$

Naive Assumption

$$P(w_1 w_2 w_3 \dots w_d | y=1)$$

$$= p(w_1 | y=1) \cdot p(w_2 | y=1) \dots p(w_d | y=1)$$

$$P(w_1 | y=1) = \frac{n_1'}{n_1}$$



$$P(y=1 | \overset{w_1 \, w_2 \dots w_d}{\text{text}}) \approx \frac{P(w_1 w_2 \dots w_d | y=1) * P(y=1)}{K}$$

$$= P(w_1 | y=1) * p(w_2 | y=1) \dots P(w_d | y=1) \cdot p(y=1)/K$$

$$\approx \prod_{i=1}^{d} p(w_i | y=1) \cdot P(y=1)/K$$

class priors

$$P(y=1 \mid text) \simeq \prod_{i=1} \underbrace{p(w_i \mid y=1)}_{\text{likelyhood}} \cdot \overbrace{P(y=1)}^{\text{class priors}} / K$$

compare

$$P(y=0 \mid text) \simeq \prod_{i=1} p(w_i \mid y=0) \cdot P(y=0) / K$$

This is Shivank

Shivank teaching python

This	Shivank	teaching	Python
1	1	0	0
0	1	1	1

Train time Complexity

Likelyhood

→ $P(y=1)$

→ $P(y=0)$

$O(nd+2)$

→ $P(w_i \mid y_i=1) \rightarrow d$

→ $O(nd)$

→ $P(w_i \mid y_i=0) \rightarrow d$

$O(nd)$

$\Downarrow$

$O(\underbrace{2nd}_{n})$

$\cancel{n^2}$

$P(w_1\ w_2 \mid y=1)$

$\approx P(w_1 \mid y=1) \cdot P(w_2 \mid y=1)$

Test time complexity

$w_1 \quad w_2 \quad w_3 \quad w_4 \quad - - - \quad w_k$

$O(k)$

Space Complexity

P(y = 0)
P(y = 1)

likelyhood : 2d

2d + 2

O(2d)
⇓
O(d)

Break : 9:16 PM

⇒ Laplace Smoothing

This is Shivank

Shivank teaching python

This	Shivank	teach	Python
1	1	0	0
0	1	1	1

Ananth teach Python

$x_q = w_1\, w_2\, w_6\, w_7\, w'$

$w' \in \{ w_1, w_2, \ldots, w_d \}$

$P(y=1 \mid w_1 w_2 w_6 w_7 w') = P(y=1) * P(w_1 \mid y=1) \ldots P(w' \mid y=1)$

$\frac{\#\text{ word } w' \text{ is part of } y=1}{\#\text{ values in } y=1} \simeq 0$

$P(y=0 \mid w_1\, w_2 \ldots w') = 0$

Laplace smoothing or Aditive smoothing

$P(w_j \mid y=1) = \frac{n_j + \alpha}{n_1 + \alpha \cdot c}$

c ← # distinct values w_j can take

100 text msg

$y = 1$

$\alpha = 1$

$C = 2$

$$P(w' \mid y=1) = \frac{0 + 1}{100 + (1*2)} = \frac{1}{102} \approx 0.0098$$

$$\Rightarrow \frac{\cancel{0} + 1}{\cancel{0} + 1*2} = \frac{1}{2}$$

Hyper Param Tuning

Train Time $\quad n_{j1} = 10 \quad n_j = 100$

Case 1

$$P(w_j \mid y=1) = \frac{n_{j1} + \alpha}{n_1 + 2\alpha} = \frac{10+1}{100+2} \approx 0.1$$

$\alpha = 1$

$$\simeq \frac{n_{j1}}{n_1} = \frac{10}{100} \approx 0.1$$

$\alpha = 10{,}000$

$$P(w_j \mid y=1) = \frac{10 + 10000}{100 + 20000} \simeq \frac{1}{2}$$

$\alpha \uparrow \rightarrow$ underfit

$P(y=1) = 0.7$

$P(y=0) = 0.3$

$P(y=1 \mid w_1 \ldots w_d) = P(y=1)\, \Pi$ ___

$P(y=0 \mid w_1 \ldots w_d) = P(y=0)\, \Pi$ ___

$\alpha \uparrow \rightarrow$ underfit

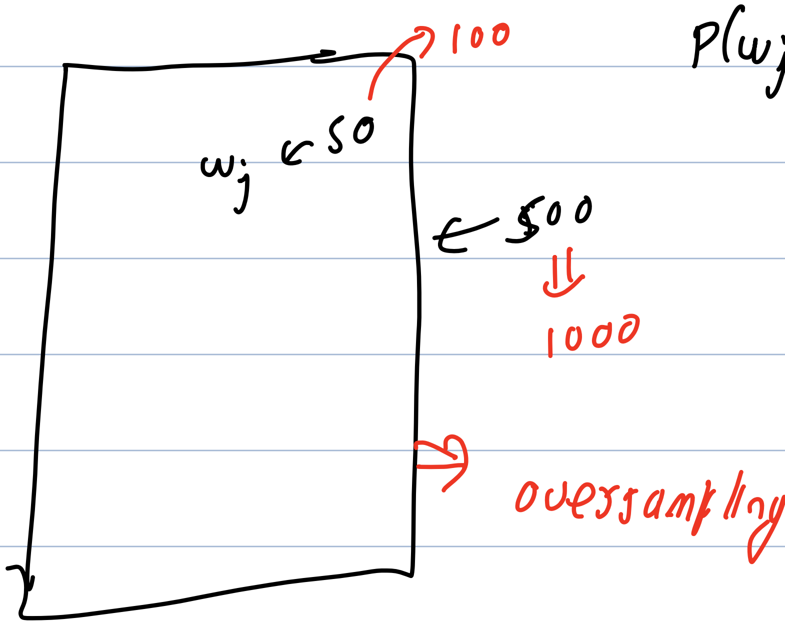
$\alpha \downarrow \longrightarrow$ overfit

$$\left.\begin{aligned} P(y=1) &= 0.7 \\ P(y=0) &= 0.3 \end{aligned}\right\} \text{ imbalance}$$

$\Downarrow$ rebalance

$$P(y=1 \mid text) \propto \underbrace{P(y=1)}_{\cancel{0.7}\ 0.5} \cdot \underbrace{\Pi \text{ likelyhood}}_{0.1} \approx 0.07$$

$$P(y=0 \mid text) \propto \underbrace{P(y=0)}_{\cancel{0.3}\ 0.6} \cdot \underbrace{\Pi \text{ likelihood}}_{0.2} \approx 0.06$$



$$P(w_j \mid y=0) = \frac{50+\alpha}{500+2\alpha} \simeq 10\%$$

$$P(w_j \mid y=1) = \frac{100+\alpha}{1000+2\alpha} \approx 10\%$$

Break: 8:12 am

$\ll 1$

$p(y=0 \mid \ldots) \simeq \ldots$

$p(y=1 \mid \ldots) \simeq \ldots$

$0.2 * 0.3 * 0.4 * 0.2 * 0.?$

$$\Rightarrow \underbrace{0.00000000023}_{\text{Floating Number}} \simeq$$

Underflow Problem

$\Rightarrow \quad \prod$

$$\log\left(\prod w_1 \cdot w_2 \cdot w_3\right) \Rightarrow \log(w_1) + \log(w_2) + \log(w_3)$$

$$\log\left(p(y=1 \mid w_1\, w_2 \ldots w_d)\right) = \log(p(y=1)) + \sum_{j=1}^{d} \log(w_j \mid y=1)$$

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

$\Rightarrow$ Feature Importance

$\boxed{p(w_1 \mid y=1)}$ $\quad$ $\boxed{p(w_2 \mid y=1)}$

$\exists$

$= w_1 w_2 \cdots \boxed{w'} \notin D_{Train}$

1

1000000000

Bernoulli NB

Multinomial NB

This is Shivank and shivank

Shivank teaching python

This	Shivank	teach	Python
1	2	0	0
0	1	1	1