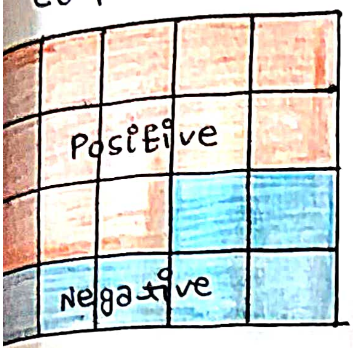
ne

abilities

's Rule (Applied in diff. field, including NLP)

your own Naive Bayes Tweet Classifier

Corpus of Tweets



Tweets containing the word "happy".

→ Positive Tweet

A) = Npos / N → 13/20

→ 0.65

egative) = 1 − P(positive)

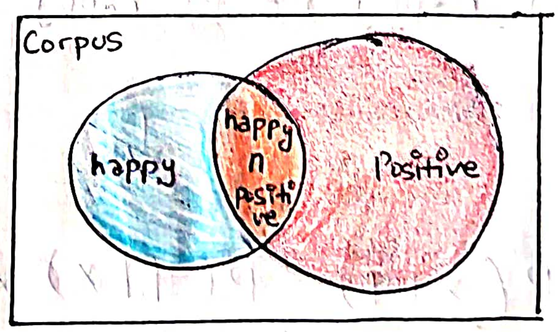
→ 1 − 0.65 → 0.35

B → tweet containing "happy"

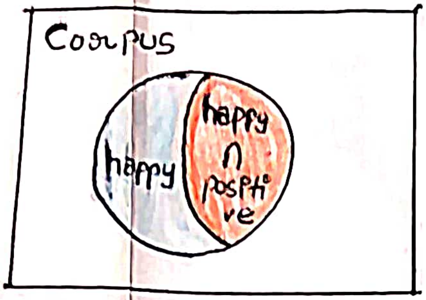
P(B) = p(happy) = Nhappy / N

P(B) = 4/20 → 0.2

obability of the intersection



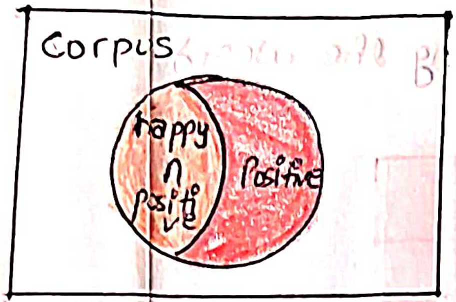
$A \cap B) = P(A, B) = \frac{3}{20} = 0.15$



$$P(A|B) = P(\text{positive} \mid \text{"happy"})$$

$$= \frac{3}{4} \Rightarrow 0.75$$

prob. of ~~tweet~~ is Positive given that it contains word happy



$$P(B|A) = P(\text{"happy"} \mid \text{positive})$$

$$= \frac{3}{13} \Rightarrow 0.231$$

Baye's Rule →

$$P(\text{Positive} \mid \text{"happy"}) = \frac{P(\text{Positive} \cap \text{"happy"})}{P(\text{"happy"})}$$

$$P(\text{"happy"} \mid \text{Positive}) = \frac{P(\text{"happy"} \cap \text{Positive})}{P(\text{Positive})}$$

$$\downarrow$$

$$P(\text{Positive} \mid \text{"happy"}) = P(\text{"happy"} \mid \text{Positive}) \times \frac{P(\text{Positive})}{P(\text{"happy"})}$$

$$\downarrow$$

$$P(x|y) = P(y|x) \times \frac{P(x)}{P(y)}$$

ive Bayes for Sentiment Analysis —

itive tweets

am happy Because I am Learning NLP.
am happy, not sad.

gative tweets

am sad, I am not learning NLP.
am sad, not happy

word
I
am
happy
because
learning
NLP
sad
not

freq

ord	Pos	neg
I	3	3
am	3	3
appy	2	1
ecause	1	0
earning	1	1
NLP	1	1
sad	1	2
not	1	2
class	13	12

ps in computing the
itional probabilities
word in the class-

word	pos	neg
I	$P\left(\frac{I}{pos}\right) = \frac{3}{13}$ $= \boxed{0.24}$	$P\left(\frac{I}{neg}\right) = \frac{3}{12}$ $\boxed{= 0.25}$
am	$P\left(\frac{am}{pos}\right) = \frac{3}{13}$ $= \boxed{0.24}$	$P\left(\frac{am}{neg}\right) = \frac{3}{12}$ $= \boxed{0.25}$
happy	0.15	0.08
because	0.08	~~0.04~~ 0
learning	0.08	0.08
NLP	0.08	0.08
sad	0.08	0.17
not	0.08	0.17
Σ	1	1

$\Downarrow$

P(Wi | class)

(4)

eg1 predict sentiment of whole tweet

Tweet : I am happy today ; I am Learning

$$\prod_{i=1}^{m} \frac{P(w_i|pos)}{P(w_i|neg)}$$

$$\frac{0.20}{0.20} * \frac{0.20}{0.20} * \frac{0.14}{0.10} * \frac{0.20}{0.20} * \frac{0.20}{0.20} * \frac{0.10}{0.10}$$

$$= \frac{0.14}{0.10} \Rightarrow 1.4 > 1$$

(positive sentiment)

a nearly identical conditional probability, words tha equally probable & don't add anything to the sent

These are your power words, These carry a lot o weights in determining in your tweet sentiments

conditional probability for the negative class is when this happens no way of comparing b/w positive tweets & negative tweets corpora.

which will become a problem for your calculations

$$\prod_{i=1}^{m} \frac{P(w_i|Pos)}{P(w_i|Neg)}$$ Naive Baye's inference conditional-rule for binary classification

This expression says you are going to take the product acr all of the words in your tweets of the probability for word in positive class divide it by the prob. in neg. cla

Bayes Classifier is a supervised learning algorithm Based on Baye's theorem.

probabilistic classifier, which mean it predict on the of probability of an Object.

theorem →

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{— ①} \qquad P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{— ②}$$

$$P(A \cap B) = P(A|B) * P(B) = P(B|A) * P(A)$$

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

or

$$P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}$$

→ Bayes Theorem

Here Y is class variable and X is dependent feature vector ~~variable~~

↳ $(x_1, x_2, x_3, \ldots, x_n)$

$$P(Y|x_1, x_2, \ldots, x_n) = \frac{P(Y) * P(x_1, x_2, x_3, \ldots, x_n | Y)}{P(x_1, x_2, x_3, \ldots, x_n)}$$

$$\rightarrow \frac{P(Y) * P(x_1|Y_1) * P(x_2|Y_2) * \ldots * P(x_n|Y_n)}{P(x_1) * P(x_2) * P(x_3) * \ldots * P(x_n)}.$$

Y: yes or no / 0 or 1 / T or F.

Day	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	~~No~~ yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Strong	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No
15	Sunny	Cool	High	Strong	?

Day 15

$$P(\underset{\substack{\uparrow \\ \text{class variable}}}{yes} \mid \underbrace{sunny, cool, high, strong}_{\text{dependent feature vector}})$$

$$\Rightarrow \frac{P(yes) * P(sunny, cool, high, strong \mid yes)}{P(sunny, cool, high, strong)}$$

$$\Rightarrow \frac{P(yes) * P(sunny \mid yes) * P(cool \mid yes, sunny) * P(high \mid yes, sunny, cool) * P(strong \mid yes, sunny, cool, \ldots)}{P(sunny, cool, high, strong)}$$

$$\rightarrow \frac{9}{14} * \frac{2/14}{9/14} * \frac{3}{9} * \frac{3}{9} * \frac{4}{9} \Rightarrow \frac{9 * 2 * 3 * \ldots}{14 * 9 * 9 * 9 \ldots}$$

$\frac{3}{9}$: $\frac{P(cool \cap yes \cap sunny)}{\cancel{P(cool)}\ P(yes \cap sunny)} \rightarrow P(cool \mid yes)$

$\frac{3}{9} \rightarrow P(high \mid yes)$

$\frac{4}{9} \rightarrow P(strong \mid yes)$

$\rightarrow 0.007$

| Sunny, cool, high, strong)

$\underbrace{\text{Sunny, cool, high, strong}}_{\text{dependent feature vector}}$

ve

No) * P(sunny | No) * P(cool | No, sunny) * P(high | ~~yes~~ No, sunny, cool)
* P(strong | No, sunny, cool, high)

(No) * P(sunny | No) * P(cool | No) * P(High | No) * P(strong | No)

$\cdot * \frac{3}{5} * \frac{1}{5} * \frac{4}{5} * \frac{3}{5} \rightarrow \frac{5 * 3 * 1 * 4 * 3}{14 * 5 * 5 * 5 * 5} \rightarrow 0.0206$

Strong → yes   P(Strong | yes) = 4/9 $\frac{P(\text{Strong} \cap \text{yes})}{P(\text{yes})}$

Strong → no   P(Strong | no) = 3/5 $\frac{P(\text{Strong} \cap \text{no})}{P(\text{no})}$

weak → yes   P(weak | yes) = 5/9

weak → no   P(weak | no) = 2/5

ity → High → yes   P(High | yes) = $\frac{P(\text{high} \cap \text{yes})}{P(\text{yes})}$ → 3/9

High → No   P(High | No) = $\frac{P(\text{High} \cap \text{No})}{P(\text{No})}$ → 4/5

Normal → yes   P(Normal | yes) = $\frac{P(\text{Normal} \cap \text{yes})}{P(\text{yes})}$ → 6/9

Normal → No   P(Normal | No) = $\frac{P(\text{Normal} \cap \text{No})}{P(\text{No})}$ → 1/5

mp

Hot → yes → 2/9
Hot → no → 2/5

Mild → yes → 4/9
Mild → no → 2/5

Cold → yes → 3/9
Cold → no → 1/5

9

Outlook → Sunny → yes → 2/9

Outlook → Sunny → no → 3/5

Outlook → Overcast → yes → 4/9

Outlook → Overcast → no → 0/5

Outlook → Rain → yes → 3/9

Outlook → Rain → no → 2/5

Laplacian Smoothing :

Sometimes you try to calculate the probability of a word happening after a word. To do that you might want to count the nos of times those two words showed up. One after another, divided by the number of times the first word appears. Now what if the ~~two~~ two words never showed up next to each other in the training corpus. You, get a probability of zero, and the probability of an entire sequence might go to zero.

Now, how can we fix this?

previously, $P(w_i|\text{class}) = \dfrac{\text{freq}(w_i,\text{class})}{N_{\text{class}}}$

$\text{class} \in \{\text{positive}, \text{negative}\}$

N_{class} = frequency of all words in class.

V_{class} = nos of unique words in class.

$$P(w_i|\text{class}) = \frac{\text{freq}(w_i,\text{class}) + 1}{N_{\text{class}} + V_{\text{class}}}$$

avoids the probo being zero.

[Laplacian Smoothing].

word	pos	neg
I	3	3
am	3	3
happy	2	1
Because	1	0
Learning	1	1
NLP	1	1
sad	1	2
not	1	2
Nclass	13	12

$\rightarrow V = 8$

word	pos	neg
I	$P(I/pos) \Rightarrow \frac{3+1}{13+8} \Rightarrow \boxed{0.19}$	$P(I\mid$ … $\Rightarrow \frac{3+}{12+}$ … $\Rightarrow \boxed{0.}$
am	0.19	0.2
happy	0.14	0.1
because	0.10	0.0
learning	0.10	0.1
NLP	0.10	0.1
sad	0.10	0.1
not	0.10	0.1
Σ	1	1

[Laplacian Smoothing]

- Laplacian smoothing to avoid $\prod_{i=1}^{m} \frac{P(w_i \mid pos)}{P(w_i \mid neg)}$

$\Downarrow$

$P(w_i \mid Class)$

$\Downarrow$

0

$$\frac{0.19}{0.20} * \frac{0.19}{0.20} * \frac{0.14}{0.10} * \frac{0.10}{0.05} * \frac{\cancel{0.10}}{\cancel{0.10}} * \frac{0.}{0.}$$

$$\Rightarrow \frac{0.000005054}{0.0000045} \Rightarrow 1.12 > 1$$