Naive Bayes

&

Support Vector Machines

[ML - 1]

Naive Bayes Recap

Conditional Probab [Baye's theorem]

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

Therefore

$$P(y=1/x) = \frac{P(x|y=1) * P(y=1)}{P(x)}$$

$$P(y=0/x) = \frac{P(x|y=0) * P(y=1)}{P(x)}$$

In Simple terms:

$$P(\text{spam}/\text{words}) = \frac{P(\text{words}/\text{spam}) * P(\text{spam})}{P(\text{words})}$$

$$P(\text{ham}/\text{words}) = \frac{P(\text{words}/\text{ham}) * P(\text{ham})}{P(\text{words})}$$

NB Classifier

$$x_i \rightarrow \begin{cases} P(\text{spam}/x_i) \\ P(\text{ham}/x_i) \end{cases} \rightarrow \; > \; \rightarrow$$

Terms

1. P(words / spam) = likelihood

2. P(words) = evidence

3. P(spam) = prior

$$\overset{\text{posterior}}{\overset{\searrow}{P(A|B)}} = \frac{\overset{\text{likelihood}}{\overset{\downarrow}{P(B|A)}} * \overset{\text{prior}}{\overset{\swarrow}{P(A)}}}{\underset{\text{evidence}}{\underset{\uparrow}{P(B)}}}$$

1. Evidence:

→ No need to calculate (common den.)

2. Prior:

→ $P(y = i)$

Easy → df.classes.value_counts(norm = True)

3. Likelihood

$P(x_i / y_i = j)$

$x_i = [w_1, w_2, \ldots, w_m]$

$$\div \; P(w_1, w_2 \ldots w_m / y = 1)$$

Naive assumption:

$$P(w_1 / y = 1) \cdot P(w_2 / y = 1) \cdots P(w_m / y = 1)$$

```
(
  df.loc[df.classes == 1]["text"].
  .contains(w1).sum() /
  (df.classess == 1).sum()
)
```

Hence,

$$p(\text{spam} / \text{words}) = \frac{p(\text{spam}) * \prod p(w_i / \text{spam})}{K}$$

Log Trick

$p(w_1 / y=1) \cdot p(w_2 / y=1) \ldots \ldots \quad (p w_{50} / y=1)$

→ These are lots of mult of values $\underline{< 1}$

Eg: $0.5^{50} \sim \underline{\underline{10^{-16}}}$:→ Possible compute issues.

hence, we use

$$\text{antilog}\left(\log\left(p(w_1/y=1)\cdot p(w_2/y=1)\cdots(p\,w_{50}/y=1)\right)\right)$$

$$= e^{\log(p_1 \cdot p_2 \cdot p \cdots p_{50})}$$

$$= e^{\log(p_1)+\log(p_2)\cdots+\log(p_{50})}$$

→ More stable computatn

Laplace Smoothing

What if we get a new word in test data?

$p(w_1/y=1) \cdot \underbrace{p(w_{new}/y=1)}_{?} \cdot - - - - \quad (p w_{50} / y=1)$

a) → 0 [Overall o/p = 0]

b) → 1 [overall o/p = unchanged]

if w_{new} is avl for $y=0$ but not $y=1$

c) → Laplace smoothing

$$p(w_i/y=1) = \frac{\# \; w_i \text{ occurs when } y=1}{\# \; y=1}$$

$$= \frac{n_i}{n}$$

$$P_{lap(k)}(w_i / y=1) = \frac{n_i + k}{n + \underbrace{c}_{\text{\# possible values of } x_i} \underbrace{k}_{\text{smoothing factor}}}$$

<u>Example</u>

H H T

		$P(H)$	$P(T)$		
No smoothing	$k=0$	$2/3$	$1/3$	66%	33%
	$k=1$	$3/5$	$2/5$	60%	40%
	$k=100$	$\frac{102}{203}$	$\frac{101}{203}$	51%	49%

So if we decide a suitable k, then the prob is not affected much.

over smooth

But, when $n_i = 0$

$$= \frac{n_i + k}{n_i + Ck} = \frac{0 + k}{n_i + Ck} \neq 0$$

small number

~ 0