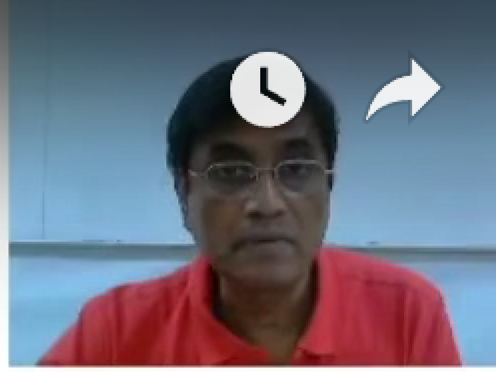
Naïve Bayes Algorithm: Multinomial Event Model

- Most commonly used model out of all three.
- Slightly different from already discussed Multi-Variate Bernoulli Event Model (Ref Lecture-24) in representing the feature vectors which are assumed to follow multinomial distribution.

- Recap:
  - GDA: A Naïve Bayes model with the assumption that the features are continuous random variables and follow Normal distribution
  - Multi Variate Bernoulli event model with the assumption that the features are binary valued and follow Bernoulli distribution
  - Multinomial event model with the assumption that the features are discrete valued and follow multinomial distribution.

$X_j \in \{1, 2, \dots V\}$, V = vocabulary.



Let us take an example: Select total 30 messages (emails/twitter/…), 20 of which are normal messages and 10 are Spam mail. Let us also consider 5 vocabulary: "Hello", "Dear", "Got", "Lottery" , "Money"

- **STEP#1:** Make a list of words with their frequencies:

$Y^{(i)}$ \ X_j, $j=1,2,\dots,5$	Hello	Dear	Got	Lottery	Money	Total words	Total Mails
$Y^{(i)}=0$ $i=1,2,\dots,20$	20	15	2	1	2	40	20
$Y^{(i)}=1$ $i=1,2,3,\dots,10$	5	2	0	8	5	20	10

Naïve Bayes Algorithm: Multinomial Event M

- **STEP # 2,** Given it is Normal (N) class, Calculate the Probability of Seeing Each word:

Calculate the probability of seeing each word if it is Normal message :

$$P(X_1 / Y^{(i)}=0) = P(\text{Hello} / \text{Normal}) = \frac{20}{40} = 0.5$$

$$P(X_2 / Y^{(i)}=0) = P(\text{Dear} / N) = \frac{15}{40} = 0.375$$

$$P(X_3 / Y^{(i)}=0) = P(\text{Got} / N) = \frac{2}{40} = 0.05$$

$$P(X_4 / Y^{(i)}=0) = P(\text{Lottery} / N) = \frac{1}{40} = 0.025$$

$$P(X_5 / Y^{(i)}=0) = P(\text{Money} / N) = \frac{2}{40} = 0.05$$

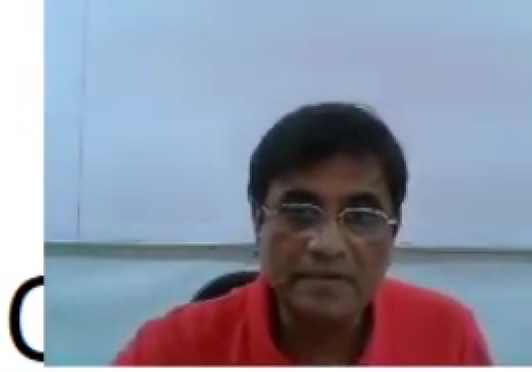
$$P(X_j / Y) \;\&\; P(Y)$$

when $j = 1, 2, \ldots 10{,}000$

$$\phi_{k|Y=0} = \frac{\sum_{i=1}^{m=20} \sum_{j=1}^{n=5} 1\{X_j^{(i)} = k \wedge Y^{(i)}=0\}}{\sum_{i=1}^{m=20} 1\{Y^{(i)}=0\} \cdot n_i}$$

$n_i \rightarrow$ no of words in the ith train mail.

Naïve Bayes Algorithm: Multinomial Event Mo



- **STEP# 3,** Given it is a Spam mail, Calculate the probability of seeing each word:

Calculate the probability of seeing each word in spam mail;

$$P(X_1 / Y^{(i)}=1) = P(\text{Hello} / \text{spam}) = 5/20 = 0.25$$

$$P(X_2 | Y^{(i)}=1) = P(\text{Dear} / \text{spam}) = \frac{2}{20} = 0.1$$

$$P(X_3 | Y^{(i)}=1) = P(\text{Got} | \text{Spam}) = 0.0$$

$$P(X_4 / Y^{(i)}=1) = P(\text{Lottery} / \text{Spam}) = \frac{8}{20} = 0.4$$

$$P(X_5 / Y^{(i)}=1) = P(\text{Money} / \text{spam}) = \frac{5}{20} = 0.25$$

In generalised form

$$\phi_{k=Y=1} = \frac{\sum_{i=1}^{m=10} \sum_{j=1}^{n=5} 1\{x_j = k\} \wedge Y^{(i)} = 1\}}{\sum_{i=1}^{m=10} 1\{Y^{(i)} = 1\} n_i}$$

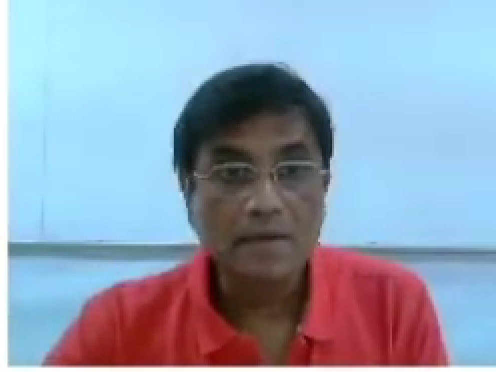
Naïve Bayes Algorithm: Multinomial Event Mo

- **STEP#4**: Calculate the class Prior Probability

$$P(Y^{(i)}=0) = \frac{20}{20+10} = 0.67 \rightarrow \phi_{Y^{(i)}=0} = \frac{\sum_{i=1}^{m=30} 1\{Y^{(i)}=0\}}{m}$$

$$P(Y^{(i)}=1) = \frac{10}{20+10} = 0.33 \rightarrow \phi_{Y^{(i)}=1} = \frac{\sum_{i=1}^{m=30} 1\{Y^{(i)}=1\}}{m}$$

$$P(x_j / Y^{(i)}) \propto P(Y^{(i)})$$



Let us take an example: Select total 30 messages (emails/twitter/...), 20 of which are normal messages and 10 are Spam mail. Let us also consider 5 vocabulary: "Hello", "Dear", "Got", "Lottery" , "Money"

List of words with their frequencies:

X_j, $j=1,2,\ldots,5$ / $Y^{(i)}$	Hello	Dear	Got	Lottery	Money	Total words	Total Mails
$Y^{(i)} = 0$ $i=1,2,\ldots,20$	20 0.5	15 0.375	2 0.05	1 0.025	2 0.05	40	20
$Y^{(i)} = 1$ $i=1,2,3,\ldots,10$	5 0.25	2 0.1	0 0.0	8 0.4	5 0.25	20	10

Naïve Bayes Classifier- Multinomial Model

Prediction :

Say now we received a message “Hello Dear”! In which class it belongs- Normal mail class or Spam class?

Let us see by calculating score of this sentence for each class:

$$P(Y^{(i)}_{=0} / X_{j,\, j=1,2}) = P(Y_0^{(i)}) \cdot \prod_{j=1}^{n=2} P(X_j / Y_0^{(i)})$$

$$P(N / \text{“Hello Dear!”}) = \overset{P(Y^{(i)}_{=0})}{0.67} \times \overset{P(\text{Hello}/N)}{0.5} \times \overset{P(\text{Dear}/N)}{0.375} = 0.126$$

$$P(N / \text{“Hello Dear!”}) \propto 0.126$$

$$P(\text{Spam} / \text{“Hello Dear!”}) = 0.33 \times 0.25 \times 0.1 = 0.0083$$

$$P(\text{Spam} / \text{“Hello Dear!”}) \propto 0.0083$$

Naïve Bayes Classifier- Multinomial Model

Prediction :

Say now we received a message "Hello Dear"! In which class it belongs- Normal mail class or Spam class?

Let us see by calculating score of this sentence for each class:

$$P(Y^{(i)}_{=0} / X_{j, j=1,2}) = P(Y_0^{(i)}) \cdot \prod_{j=1}^{n=2} P(X_j / Y_0^{(i)})$$

$P(Y^{(i)}_{=0})$ $\quad$ $P(\text{Hello}/N)$ $\quad$ $P(\text{Dear}/N)$

$$P(N(\text{"Hello Dear!"})) = 0.67 \times 0.5 \times 0.375 = 0.126$$

$$P(N / \text{"Hello Dear!"}) \propto 0.126$$

$$P(\text{Spam} / \text{"Hello Dear!"}) = 0.33 \times 0.25 \times 0.1 = 0.0083$$

$$P(\text{Spam} / \text{"Hello Dear!"}) \propto 0.0083$$

$$\arg\max_{Y} P(Y^{(i)} / X_j) \rightarrow Y = 0$$

Naïve Bayes Classifier- Multinomial Mode

Now let us assume another mail has arrived to your mailbox and you need to classify: " Hello ! Got Lottery Money"!

1. Score of the sentence in the normal class:

$$P(Y^{(i)}_{=0}) \cdot \prod_{j=1}^{n=4} P(X_j / Y^{(i)}_{=0})$$

$$= 0.67 \times P(\text{Hello}/N) \times P(\text{Got}/N) \times P(\text{Lottery}/N) \times P(\text{Money}/N)$$

$$= 0.67 \times 0.5 \times 0.05 \times 0.025 \times 0.05$$

$$= 0.000021$$

$$P(N / \text{"Hello ! Got Lottery Money !"}) \propto 0.000021$$

$$P(\text{Spam} / \text{"Hello ! Got Lottery Money !"}) \rightarrow 0.33 \times 0.25 \times 0.0 \times 0.4 \times 0.25$$

$$= 0$$

Let us take an example: Select total 30 messages (emails/twitter/…), 20 of which are normal messages and 10 are Spam mail. Let us also consider 5 vocabulary: "Hello", "Dear", "Got", "Lottery" , "Money"

List of words with their frequencies:

X_j, $j=1,2,\dots,5$ / $Y^{(i)}$	Hello	Dear	Got	Lottery	Money	Total words	Total Mails
$Y^{(i)}=0$ $i=1,2,\dots,20$	20 $+\alpha$ ($\alpha = 1$) 0.5	15 $+\alpha$ ($\alpha = 1$) 0.375	2 $+\alpha$ ($\alpha = 1$) 0.05	1 $+\alpha$ ($\alpha = 1$) 0.025	2 $+\alpha = 1$ 0.05	40 $+5\alpha = 1$ 45	20
$Y^{(i)}=1$ $i=1,2,3,\dots,10$	5 $+\alpha$ ($\alpha = 1$) 0.25	2 $+\alpha$ ($\alpha = 1$) 0.1	0 $+\alpha$ ($\alpha = 1$) 0.0	8 $+\alpha$ ($\alpha = 1$) 0.4	5 $+\alpha = 1$ 0.25	20 $+5\alpha = 1$ 25	10

$\alpha = 1$

Naïve Bayes Classifier- Multinomial Mode

Given it is a Normal class recalculate the probability of seeing each word :

(1) For normal class (N) $\longrightarrow$

$$P(X_1 | Y^{(i)}_{=0}) = P(\text{Hello} | N) = \frac{21}{45} = 0.47$$

$$P(X_2 | Y^{(i)}_{=0}) = P(\text{Dear} | N) = \frac{16}{45} = 0.356$$

$$P(X_3 / Y^{(i)}_{=0}) = P(\text{Got} / N) = \frac{3}{45} = 0.067$$

$$P(X_4 / Y^{(i)}_{=0}) = P(\text{Lottery} / N) = \frac{2}{45} = 0.04$$

$$P(X_5 | Y^{(i)}_{=0}) = P(\text{Money} / N) = \frac{3}{45} = 0.067$$

$$\phi_{k/Y^{(i)}=0} = \frac{\sum_{i=1}^{m=20} \sum_{j=1}^{n=5} 1\{X_j^{(i)} = k \wedge Y^{(i)}_{=0}\} + \alpha \ (\text{say} = 1)}{\sum_{i=1}^{m=20} 1\{Y = 0\} n_i + 5\alpha \ (\alpha = 1)}$$

Naïve Bayes Classifier- Multinomial Mode

Given it a Spam class recalculate the probability of each word:

$$P(X_1 | Y^{(i)}=1) = P(\text{Hell}/\text{spam}) = \frac{6}{25} = 0.24$$

$$P(X_2 | Y^{(i)}=1) = P(\text{Dear}/\text{spam}) = \frac{3}{25} = 0.12$$

$$P(X_3 | Y^{(i)}=1) = P(\text{Got}/\text{spam}) = \frac{1}{25} = 0.04$$

$$P(X_4 / Y^{(i)}=1) = P(\text{Lottery}/\text{spam}) = \frac{9}{25} = 0.36$$

$$P(X_5 / Y^{(i)}=1) = P(\text{Money}/\text{spam}) = \frac{6}{25} = 0.24$$

$$\Phi_{k/Y=1} = \frac{\sum_{i=1}^{m=10} \sum_{j=1}^{n=5} 1\{x_j = k \wedge Y^{(i)}=1\} + \underset{=1}{\alpha}}{\sum_{i=1}^{m=10} 1\{Y^{(i)}=1\} n_i + |V|} \quad (\text{Say } \alpha = 1)$$

Naïve Bayes Classifier- Multinomial Mode

New Prediction after Laplace smoothing:

New message/mail — "Hello! Got Lottery Money!"

$P(N / \text{"Hello! Got Lottery Money!"})$

$\Rightarrow P(Y^{(i)}=0) \prod_{j=1}^{4} P(X_j^{(i)} / Y^{(i)}=0)$

$\Rightarrow 0.67 \times P(\text{Hello}/N) \times P(\text{Got} | N) \times P(\text{Lottery}/N) \times P(\text{Money}/N)$

$= 0.67 \times 0.47 \times 0.067 \times 0.04 \times 0.067 = 0.000057$

$P(N / \text{Hello! Got Lottery Money}) \propto 0.000057$

$P(\text{SPAM} / \text{"Hello! Got Lottery Money"}) \Rightarrow P(Y^{(i)}=1) \prod_{j=1}^{4} P(X_j^{(i)} / Y^{(i)}=1)$

$\Rightarrow 0.33 \times P(\text{Hell}/\text{spam}) \times P(\text{Got}/\text{spam}) \times P(\text{Lottery}/\text{spam}) \times P(\text{Money}/\text{spam})$

$\Rightarrow 0.33 \times 0.24 \times 0.04 \times 0.36 \times 0.24 = 0.00027$

$\arg\max_{Y} P(Y/X) \Rightarrow Y=1 \rightarrow$ The message belongs to SPAM

Applications of Naïve Bayes Classifier

- Text Classification
- Spam Filtering
- Sentiment analysis (positive/negative customer sentiment)
- Fraud detection
- Recommendation system
- Multi Class prediction
- Real time prediction (since it is very fast)
- Highly scalable & efficient

Limitations:

- It is very quick but rough
- Being Naïve , require to add some module to make it efficient (e.g word embedding)