# eMasters in Communication Systems

## Prof. Aditya Jagannatham

# Elective Module:

## Advanced ML Techniques

# Chapter 5

# Naïve Bayes

Bayesian Technique.

- Prior
- Posterior

# Naïve Bayes

- ***Naïve Bayes*** is best suited for ML applications… Binary Response $y \in \{0, 1\}$
  - wherein the feature vectors $\bar{\mathbf{x}}$ are Discrete

| | $y$ | $\bar{x}$ |
|---|---|---|
| Linear regression | continuos | continuos |
| Logistic Regression | Discrete | continuos |
| Naive Bayes | Discrete | Discrete |

# Naïve Bayes

- ***Naïve Bayes*** is best suited for ML applications…
  - wherein the feature vectors $\bar{\mathbf{x}}$ are **discrete**

## Naïve Bayes

$y = 1$ Spam
$y = 0$ Genuine.

- Example: ML-based e-mail **spam filter**
- **Like/ dislike** an item based on other preferences

movie

# Naïve Bayes — Email classification

Spam or Genuine.

- Consider a **feature vector** $\bar{\mathbf{x}}$ of size $N$
  - where $N$ is the number of words in the English Dictionary

$N$

~100,000 words.

each entry corresponds to a word.

$$\bar{x} = \left[\quad\right] \Big\} \approx 100{,}000$$

# Naïve Bayes

- Consider a **feature vector** $\bar{\mathbf{x}}$ of size $N$
  - where $N$ is the number of words in the **English language dictionary**

# Naïve Bayes

- The labels $y = 1,0$ indicate SPAM, GENUINE e-mails, respectively

# Naïve Bayes

- The labels $y = 1{,}0$ indicate ***spam*, *genuine*** e-mails, respectively

# Naïve Bayes

- The entry $x_j = 1$, if the ***email*** contains the $j$ –th word of the ***dictionary***,…
  - else $x_j = 0$

Each j corresponds to a word in dictionary

# Naïve Bayes

Feature vector

$$\bar{\mathbf{x}} = \begin{bmatrix} \vdots \\ 1 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix} \begin{array}{ll} \vdots & \\ \text{able} & 4562 \\ \text{above} & 4563 \\ \text{abroad} & 4564 \\ \text{access} & 4565 \\ \vdots & \vdots \end{array}$$

input vector

N = # Features.

$x_{4562} = 1$

$x_{4563} = 0$

$x_{4564} = 0$

$x_{4565} = 1$

# Naïve Bayes

Simple, NOT complicated.

- **Naïve Bayes assumption**:
- The different words are conditionally independent given the label $y$

Derogatory

# Naïve Bayes

- **Naïve Bayes assumption**:
- The different words are **conditionally independent** given the label $y$

Simplistic assumption

Reasonably True

# Naïve Bayes

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = u)$$

$u \in \{0, 1\}$

$V_i \in \{0, 1\}$

$$= p(x_1 = v_1, x_2 = v_2, \ldots, x_N = v_N \mid y = u)$$

$$= p(x_1 = v_1 \mid y = u) \times \cdots \times p(x_N = v_N \mid y = u)$$

$$= \prod_{j=1}^{N} p(x_j = v_j \mid y = u)$$

# Naïve Bayes Assumption

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = u)$$

$$= p(x_1 = v_1, \ldots, x_N = v_N | y = u)$$

$$= p(x_1 = v_1 | y = u) \times \cdots \times p(x_N = v_N | y = u)$$

$$= \prod_{j=1}^{N} p(x_j = v_j | y = u)$$

conditional independence

conditionally independent given response.

# Naïve Bayes

- The quantities $p(x_j = v_j | y = u)$ are the prior probabilities.
  - How to calculate these?

Prior

Posterior

# Naïve Bayes

$j = 1, 2, \ldots, N$

- The quantities $p(x_j = v_j | y = u)$, are the **prior probabilities**
  - How to calculate these?

$p(x_{4562} = 1 | y = 1)$

What is probability word able occurs in a spam email.

$p(x_{4562} = 1 | y = 0)$

What is the probability word able occurs in a genuine email.

Probabilities that $j^{th}$ word occurs in genuine/spam emails.

# Naïve Bayes

Labels.

- Consider the availability of $M$ ***training pairs*** $\left(\bar{\mathbf{x}}(i), y(i)\right)$

$i = 1, 2, \ldots, M$

M training points.
M total emails.
some spam
some genuine.

# Naïve Bayes

- The various ***prior probabilities*** can now be calculated as follows

# Naïve Bayes

$$1(x) = \begin{cases} 1 & \text{if } x \text{ is true} \\ 0 & \text{else.} \end{cases}$$

$$p(x_j = 1 | y = 1) = \frac{\text{number of spam emails with } j^{th} \text{ word}}{\text{Total Number of spam emails.}}$$

Prior probability.

$j^{th}$ word occurs in spam email.

$$= \frac{\underbrace{\sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 1)}_{\text{Total number of spam emails with } j^{th} \text{ word.}}}{\underbrace{\sum_{i=1}^{M} 1(y(i) = 1)}_{\text{total number of spam emails.}}}$$

# Naïve Bayes

$$p(x_j = 1 | y = 0) = \frac{\text{Number of genuine emails with } j^{th} \text{ word.}}{\text{Total number of genuine emails.}}$$

$$= \frac{\sum_{i=1}^{M} 1(x_j(i)=1, y(i)=0)}{\sum_{i=1}^{M} 1(y(i)=0)}$$

Prior probability $j^{th}$ word occurs in a genuine email.

# Naïve Bayes

$$p(x_j = 1|y = 1) = \frac{\sum_{i=1}^{M} \mathbf{1}(x_j(i) = 1, y(i) = 1)}{\sum_{i=1}^{M} 1(y(i) = 1)}$$

$$p(x_j = 1|y = 0) = \frac{\sum_{i=1}^{M} \mathbf{1}(x_j(i) = 1, y(i) = 0)}{\sum_{i=1}^{M} 1(y(i) = 0)}$$

Prior probabilities.

# Naïve Bayes

$$p(y = 1) = \frac{\text{Number of Spam emails}}{\text{Total number of emails}}$$

$$= \frac{\sum_{i=1}^{M} 1(y(i) = 1)}{M}$$

Prior probability of spam email

# Naïve Bayes

$$p(y=1) = \frac{\sum_{i=1}^{M} \mathbf{1}(y(i)=1)}{M}$$

Prior probability IF getting a spam email.

# Naïve Bayes

$$p(x_j = 0|y = 1) = 1 - p(x_j = 1|y = 1).$$

$$p(x_j = 0|y = 0) = 1 - p(x_j = 1|y = 0)$$

Prior prob. jth word does NOT occurs in genuine = 1 − prob. jth word occurs in genuine.

Prior prob. jth word does NOT occur in spam = 1 − prob. jth word occurs in spam.

# Naïve Bayes

$$p(x_j = 0|y = 1) = 1 - p(x_j = 1|y = 1)$$

$$p(x_j = 0|y = 0) = 1 - p(x_j = 1|y = 0)$$

Prior probabilities of NON occurrence.

# Naïve Bayes

$$p(y = 0) = 1 - p(y=1).$$

Prob. of genuine email
= 1 - Prob. spam. email.

# Naïve Bayes

$$p(y = 0) = 1 - p(y = 1)$$

# Naïve Bayes

- Finally, the **posterior probabilities** are calculated as follows

# Naïve Bayes

$$p(y = 1|\bar{\mathbf{x}} = \bar{\mathbf{v}}) = \frac{p(\bar{x} = \bar{v} \mid y = 1)\, p(y = 1)}{p(\bar{x} = \bar{v})}.$$

$$p(y = 0|\bar{\mathbf{x}} = \bar{\mathbf{v}}) = \frac{p(\bar{x} = \bar{v} \mid y = 0)\, p(y = 0)}{p(\bar{x} = \bar{v})}$$

Follow from Bayes rule.

$x_j = 1$ if $j^{th}$ word occurs in email

$x_j = 0$ else

# Naïve Bayes

$$p(y=1|\bar{\mathbf{x}}=\bar{\mathbf{v}}) = \frac{p(\bar{\mathbf{x}}=\bar{\mathbf{v}}|y=1) \times p(y=1)}{p(\bar{\mathbf{x}}=\bar{\mathbf{v}})}$$

$$p(y=0|\bar{\mathbf{x}}=\bar{\mathbf{v}}) = \frac{p(\bar{\mathbf{x}}=\bar{\mathbf{v}}|y=0) \times p(y=0)}{p(\bar{\mathbf{x}}=\bar{\mathbf{v}})}$$

# Naïve Bayes

- E-mail is classified as ***spam*** if

$$\underbrace{p(y=1 \mid \bar{x}=\bar{v})}_{\text{Probability spam given } \bar{x}} > \underbrace{p(y=0 \mid \bar{x}=\bar{v})}_{\text{Probability genuine given } \bar{x}}.$$

# Naïve Bayes

- E-mail is classified as ***spam*** if
  $p(y = 1|\bar{\mathbf{x}} = \bar{\mathbf{v}}) > p(y = 0|\bar{\mathbf{x}} = \bar{\mathbf{v}})$

# Naïve Bayes

$$p(y=1|\bar{\mathbf{x}}=\bar{\mathbf{v}}) > p(y=0|\bar{\mathbf{x}}=\bar{\mathbf{v}})$$

$$\Rightarrow \frac{p(\bar{x}=\bar{v}|y=1)\,p(y=1)}{\cancel{p(\bar{x}=\bar{v})}} > \frac{p(\bar{x}=\bar{v}|y=0)\,p(y=0)}{\cancel{p(\bar{x}=\bar{v})}}$$

condition for spam

$$\Rightarrow p(\bar{x}=\bar{v}|y=1)\,p(y=1) > p(\bar{x}=\bar{v}|y=0)\,p(y=0)$$

$$p(\bar{x}=\bar{v}|y=1)\,p(y=1) < p(\bar{x}=\bar{v}|y=0)\,p(y=0)$$

condition for genuine.

# Naïve Bayes

$$p(y = 1|\bar{\mathbf{x}} = \bar{\mathbf{v}}) > p(y = 0|\bar{\mathbf{x}} = \bar{\mathbf{v}})$$

$$\Rightarrow \frac{p(\bar{\mathbf{x}} = \bar{\mathbf{v}}|y = 1) \times p(y = 1)}{p(\bar{\mathbf{x}} = \bar{\mathbf{v}})} > \frac{p(\bar{\mathbf{x}} = \bar{\mathbf{v}}|y = 0) \times p(y = 0)}{p(\bar{\mathbf{x}} = \bar{\mathbf{v}})}$$

$$\Rightarrow p(\bar{\mathbf{x}} = \bar{\mathbf{v}}|y = 1) \times p(y = 1) > p(\bar{\mathbf{x}} = \bar{\mathbf{v}}|y = 0) \times p(y = 0)$$

condition for spam.

# Naïve Bayes

$$\left(\prod_{j=1}^{N} p(x_j = v_j \mid y=1)\right) p(y=1) > \left(\prod_{j=1}^{N} p(x_j = v_j \mid y=0)\right) p(y=0)$$

# Naïve Bayes

- This implies

$$\underbrace{\prod_{j=1}^{N} p\left(x_j = v_j | y = 1\right) \times p(y = 1)}_{Q_1} > \underbrace{\prod_{j=1}^{N} p\left(x_j = v_j | y = 0\right) \times p(y = 0)}_{Q_0}$$

# Naïve Bayes

$p(y=1 \mid \bar{x}=\bar{v})$. $p(y=0 \mid \bar{x}=\bar{v})$.

- Taking the logarithm yields

log aposteriori probability

classified as spam if

$$\left(\sum_{j=1}^{N} \ln p(x_j=v_j \mid y=1)\right) + \ln p(y=1)$$

$$> \left(\sum_{j=1}^{N} \ln p(x_j=v_j \mid y=0)\right) + \ln p(y=0)$$

# Naïve Bayes

$$\Rightarrow \sum_{j=1}^{N} \ln\left(p(x_j = v_j | y = 1)\right) + \ln p(y = 1)$$

$$> \sum_{j=1}^{N} \ln\left(p(x_j = v_j | y = 0)\right) + \ln p(y = 0)$$

# Laplace Smoothing

- **Naïve Bayes** has a problem

# Laplace Smoothing

- Lets say a ***new word*** "IITK" appears in your e-mail
  - which is **NOT present** in any training e-mails

NOT present in any genuine or spam email.

# Laplace Smoothing

- Lets say ***index*** of "IITK" in the ***dictionary*** is $j$.

25,634.

# Laplace Smoothing

Prior prob. of IITK in Spam.

- The **prior probabilities** will be

$$p(x_j = 1|y = 1) = \frac{\text{Number of spam emails containing IITK}}{\text{Total number of spam emails}} = 0$$

= 0

$$p(x_j = 1|y = 0) = \frac{\sum_{i=1}^{M} 1(x_j(i)=1, y(i)=0)}{\sum_{i=1}^{M} 1(y(i)=0)}.$$

0.

$$= \frac{\text{Number of genuine emails with IITK}}{\text{Total number of genuine emails}} = 0$$

Prior prob of IITK in genuine

$\Rightarrow$ Posterior prob $= 0$.

# Laplace Smoothing

- The **prior probabilities** will be

$$p(x_j = 1|y = 1) = \frac{\sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 1)}{\sum_{i=1}^{M} 1(y(i) = 1)} = 0$$

$$p(x_j = 1|y = 0) = \frac{\sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 0)}{\sum_{i=1}^{M} 1(y(i) = 0)} = 0$$

Prior prob = 0

$\Rightarrow$ Posterior prob = 0 $\rightarrow$ This causes problems in classification

## Laplace Smoothing

- These cause problems in computation of the **posterior probabilities**.

# Laplace Smoothing

- Therefore, we use the following ***prior probabilities*** instead

$$p(x_j = 1|y = 1) = \frac{1 + \sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 1)}{2 + \sum_{i=1}^{M} 1(y(i) = 1)}$$

# Laplace Smoothing

$$p(x_j = 1 | y = 0) = \frac{1 + \sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 0)}{2 + \sum_{i=1}^{M} 1(y(i) = 0)}$$

Because of addition of 1,
Prior probabilities are
never 0.

## Laplace Smoothing

- Therefore, we use the following ***prior probabilities*** instead

$$p(x_j = 1|y = 1) = \frac{1 + \sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 1)}{2 + \sum_{i=1}^{M} 1(y(i) = 1)}$$

$$p(x_j = 1|y = 0) = \frac{1 + \sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 0)}{2 + \sum_{i=1}^{M} 1(y(i) = 0)}$$

LAPLACE SMOOTHING!

# Laplace Smoothing

- This is termed ***Laplace smoothing***

Avoids prior probabilities = 0.

Equiv to adding 4 training set:

2 Spam emails
- one containing all words in dictionary
- one blank

2 Genuine emails.
- one containing all words.
- one blank.

Instructors may use this white area (14.5 cm / 25.4 cm) for the text.
Three options provided below for the font size.

Font: Avenir (Book), Size: 32, Colour: Dark Grey

Font: Avenir (Book), Size: 28, Colour: Dark Grey

Font: Avenir (Book), Size: 24, Colour: Dark Grey

Do not use the space below.