Practice assignment

Question 1

Statement

Consider a 3-class classification dataset with labels $0, 1,$ and 2. The data points belong to $\{0,1,2\}^3$. If we apply the generative model-based algorithm on the same dataset, how many features need to be estimated? Assume that the features given the label are not independent.

Answer

80

Solution

We need to estimate the parameters for $P(y=0), P(y=1)$. Since $P(y=0)+P(y=1)+P(y=2)=1$, we need to estimate two parameters for the distribution of y.

For the distribution of $x|y=0$, we need to estimate the parameters for $P(x|y=0)$.

Since $x \in \{0,1,2\}^3$, we can have $3^3 = 27$ possible data points and we need to estimate the probability for 26 such points as the sum will be one.

Similarly, for $x|y=1$ and $x|y=2$, we need 26 parameters each.

Therefore, total parameters to estimate = $2+3(26)=80$

Question 2

In question 1, if the features are conditionally independent given the labels, how many parameters need to be estimated?

Answer

20

Solution

We need to estimate the parameters for $P(y=0), P(y=1)$. Since $P(y=0)+P(y=1)+P(y=2)=1$, we need to estimate two parameters for the distribution of y.

For a given label (say $y=0$), we need to estimate
$P(f_1=0|y=0), P(f_1=1|y=0), P(f_2=0|y=0), P(f_2=1|y=0), P(f_3=0|y=0), P(f_3=1|y=0)$)

Similarly, for labels $y=1$ and $y=2$.

Therefore, total parameters to estimate = $2+3(6)=20$

Common data for questions 3, 4, and 5

Statement

Consider a naive Bayes model is trained on the following data matrix X of shape (d,n) and corresponding label vector y:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad y = [1,1,0]^T$$

Assume that $\hat{p}$ and $\hat{p}_j^{y_i}$ are estimates for $P(y=1)$ and $P(f_j=1|y=y_i)$, respectively. Here, $f_i;\ i=1,2$ is the i^{th} feature. These parameters are estimated using MLE.

Question 3

Statement

If a test point has label 0, what will be the probability that the point is $[0,0]^T$?

Options

(a)

$\hat{p}_1^0 \times \hat{p}_2^0 \times (1-\hat{p})$

(b)

$(1-\hat{p}_1^0) \times (1-\hat{p}_2^0) \times (1-\hat{p})$

(c)

$\hat{p}_1^0 \times \hat{p}_2^0$

(d)

$(1-\hat{p}_1^0) \times (1-\hat{p}_2^0)$

Answer

(d)

Solution

We know that $\hat{p}_j^{y_i}$ is the estimate for $P(f_j = 1|y = y_i)$. It implies that $(1-\hat{p}_j^{y_i})$ is the estimate for $P(f_j = 0|y = y_i)$

Therefore,

$$\begin{aligned} P(x = [0,0]^T|y = 0) &= P(f_1 = 0|y = 0).\, P(f_2 = 0|y = 0) \\ &= (1-\hat{p}_1^0) \times (1-\hat{p}_2^0) \end{aligned}$$

Question 4

Statement

What is the value of p_1^1?

Answer

1

Solution

$\hat{p}_1^1$ is the estimate for $P(f_1 = 1|y = 1)$

$$\hat{p}_1^1 = \frac{\sum\limits_{i=1}^{n} \mathbb{1}(f_1 = 1, y = 1)}{\sum\limits_{i=1}^{n} \mathbb{1}(y = 1)}$$

Here, the first two examples belong to label 1 and the first feature value for both examples is 1, therefore

$\hat{p}_1^1 = 1$

Question 5

Statement

What will be the probability that a test data point $[0,1]$ is labeled as 0? Assume no smoothing of data is done.

Answer

0

Solution

$$P(y=0|x=[0,1]) = \frac{P(x=[0,1]|y=0).\,P(y=0)}{P(x=[0,1])}$$
$$= \frac{(1-\hat{p}_1^0)\hat{p}_2^0\hat{p}}{P(x=[0,1])}$$

Here

$\hat{p}_2^0 = 0$ since, label zero example takes only zeros for all the features

Therefore,

$P(y=0|x=[0,1]) = 0$

Question 6

Statement

Consider a spam classification problem that was modeled using naive Bayes. The features take a value of 1 or 0 depending on whether a word is present in the email or not. Assume that the probability of a mail being spam is 0.2. The following table gives the estimation for conditional probabilities for some of the words:

word	**label**	$P(\text{word}|\text{label})$
Hurray!	spam	0.7
win	spam	0.2
exciting	spam	0.01
prizes	spam	0.3
Hurray!	Non-spam	0.01
win	Non-spam	0.02
exciting	Non-spam	0.01
prizes	Non-spam	0.1

Consider a mail with the following sentence: "Hurray! win exciting prizes"

With what probability the mail will be predicted spam? Assume that these are the only possible words (that is there are four features) in a mail. Write your answer correct to two decimal places. A

Answer

0.99 Range: $[0.98. 1]$

Solution

$$P(y=\text{spam}|\text{mail}) = \frac{P(\text{mail}|\text{spam}).\,P(\text{spam})}{P(\text{mail}|\text{spam}).\,P(\text{spam}) + P(\text{mail}|\text{non-spam}).\,P(\text{non-spam})}$$

Here,

$P(\text{spam}) = 0.2, \quad P(\text{non-spam}) = 0.8$

Denote spam as 0 and non-spam as 1.

Therefore,

$$\begin{aligned}
P(y=0|\text{mail}) &= \frac{P(\text{mail}|0)(0.2)}{P(\text{mail}|0)(0.2) + P(\text{mail}|1)(0.8)} \\
&= \frac{P(\text{Hurray!}|0)P(\text{win}|0)P(\text{exciting}|0)P(\text{prizes}|0)(0.2)}{P(\text{Hurray!}|0)P(\text{win}|0)P(\text{exciting}|0)P(\text{prizes}|0)(0.2) + P(\text{Hurray!}|1)P(\text{win}|1)P(\text{exciting}|1)P(\text{prizes}|1)(0.8)} \\
&= \frac{0.7(0.2)(0.01)(0.3)(0.2)}{0.7(0.2)(0.01)(0.3)(0.2) + 0.01(0.02)(0.01)(0.1)(0.8)} \\
&= 0.99
\end{aligned}$$

Question 7

Statement

A binary classification dataset contains only one feature and the data points given the label follow the gaussian distributions whose means and variances are already estimated as:

$$x|(y=0) \sim \mathrm{N}(0,1)$$
$$x|(y=1) \sim \mathrm{N}(2,2)$$

What will be the decision boundary learned using the naive Bayes algorithm? Assume that $\hat{p}$, an estimate for $P(y=1)$, is 0.5.

Hint: Solve $P(y=1|x) = P(y=0|x)$

Options

(a)

$$\frac{x^2}{2} - \frac{(x-2)^2}{4} = \frac{1}{2}\ln 2$$

(b)

$$\frac{x^2}{4} = \frac{1}{2}\ln 2$$

(c)

$$4x = 2\ln 2 + 4$$

(d)

$$\frac{x^2}{4} - \frac{(x-2)^2}{2} = \ln 2$$

Answer

(a)

Solution

The decision boundary is given by

$$\{x : P(y=1|x) = P(y=0|x)\}$$

$$\begin{aligned}
&P(y=1|x) = P(y=0|x)\\
\Rightarrow &\frac{P(x|y=1).\,P(y=1)}{P(x)} = \frac{P(x|y=0).\,P(y=0)}{P(x)}\\
\Rightarrow &P(x|y=1) = P(x|y=0) \quad (\because P(y=0) = P(y=1) = 0.5)\\
\Rightarrow &\frac{1}{\sqrt{2\pi}\sqrt{2}}\exp(-(x-2)^2/4) = \frac{1}{\sqrt{2\pi}}\exp(-(x)^2/2)\\
\Rightarrow &\exp(-(x-2)^2/4) = \sqrt{2}\exp(-(x)^2/2)\\
\Rightarrow &\ln(\exp(-(x-2)^2/4)) = \ln(\sqrt{2}\exp(-(x)^2/2))\\
\Rightarrow &\frac{-(x-2)^2}{4} = \frac{1}{2}\ln 2 + \frac{-x^2}{2}\\
\Rightarrow &\frac{-x^2}{2} - \frac{-(x-2)^2}{4} = \frac{1}{2}\ln 2
\end{aligned}$$

Common data for questions 8, 9 and 10

Statement

Consider the gaussian naive Bayes algorithm was run on the following dataset:

feature 1 (f_1)	**feature 2** (f_2)	**Label**
1.5	1.6	0
2.1	2.4	1

feature 1 (f_1)	feature 2 (f_2)	Label
2.9	1.5	1
1.7	0.8	1

Question 8

Statement

What will be the value of $\hat{p}$?

Answer

0.75 Range; [0.74, 0.76]

Solution

$$\hat{p} = \frac{\sum\limits_{i=1}^{n} y_i}{n} = \frac{3}{4}$$

Question 9

Statement

What will be the value of $\hat{\mu}_0$?

Options

(a)

1.5

(b)

1.55

(c)

$(1.5, 1.6)$

(d)

$(2.23, 1.56)$

Answer

(c)

Solution

$$\begin{aligned}\hat{\mu}_0 &= \frac{\sum\limits_{i=1}^{n} \mathbb{1}(y_i = 0)x_i}{\sum\limits_{i=1}^{n} \mathbb{1}(y_i = 0)} \\ &= \frac{(1.5, 1.6)}{1}\end{aligned}$$

Question 10

Statement

What will be the value of $\hat{\mu}_1$?

Options

(a)

2.23

(b)

1.56

(c)

$(1.5, 1.6)$

(d)

$(2.23, 1.56)$

Answer

(d)

Solution

$$\begin{aligned}\hat{\mu}_1 &= \frac{\sum\limits_{i=1}^{n} \mathbb{1}(y_i = 1)x_i}{\sum\limits_{i=1}^{n} \mathbb{1}(y_i = 1)} \\ &= \frac{[2.1, 2.4] + [2.9, 1.5] + [1.7, 0.8]}{3} \\ &= (2.23, 1.56)\end{aligned}$$