Week_8_mlt_questions

2024 Mar 24 Quiz 2

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

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

Assume that $\hat{p}$ and $\hat{p}_j^{y_i}$ are estimates for $P(y = 1)$ and $P(f_j = 1 \mid y = y_i)$, respectively. Here, f_i; $i = 1, 2, 3$ is the i^{th} feature. These parameters are estimated using MLE.
Based on the above data, answer the given subquestions.

i. Calculate the value of $\hat{p}_2^1$

[Ans: 0.30 to 0.35]

ii. Calculate the value of $\hat{p}_1^1$

[Ans: 0.30 to 0.35]

2. Suppose you have a five-class classification problem where class label $y \in \{0, 1, 2, 3, 4\}$ and each training example x_i has binary features $f_1, f_2, f_3 \in \{0, 1\}$. How many parameters do we need to know to classify an example using Naive Bayes classifier?

[Ans: 19 to 20]

2024 Apr 28 ET

3. Consider the following dataset with 6 samples along with the corresponding labels. Each sample has three binary features f_1, f_2 and f_3.

sample	f_1	f_2	f_3	y
x_1	1	1	0	1
x_2	0	1	0	1
x_3	1	1	1	0
x_4	0	1	1	0
x_5	1	0	1	0
x_6	1	1	1	1

Assume that the features are conditionally independent given the label y. Suppose the test sample is $x_{\text{test}} = [0, 1, 0]^T$.
What is the estimated probability that the test point belongs to class 0 (that is, $p(y = 0 \mid x_{\text{test}})$)?

[Ans: 0]

2024 Aug 04 Quiz 2

4. **Comprehension:**
Assume that there is no smoothing. Answer the given subquestions.
Consider the following dataset for a binary classification problem in which the features are from $\{0, 1\}^3$ and the labels are from $\{0, 1\}$.

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

A Naive Bayes classifier is trained on this dataset. The parameters to be estimated are represented as $\hat{p}_i^y$, which are presented in the form of the table given below. Recall that $\hat{p}_i^y$ is the probability of feature i taking a value 1 in class y:

i	$y = 0$	$y = 1$
1	a	b
2	c	d
3	e	f

i. Find the value of $c + d$.

[Ans: 1.25]

ii. Find the value of $a + c + e$.

[Ans: 1.5]

iii. Find the predicted label for the test point $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

[Ans: 0]

5. **Comprehension:**
Consider a multi-class classification problem with 3 classes and 4 features, all of which are binary. A generative model is used to model the joint distribution of the features and labels. Based on the above data, answer the given subquestions.

i. Find the total number of independent (free) parameters in the model if the class conditional independence assumption is not enforced.

[Ans: 47]

ii. Find the total number of independent (free) parameters in the model if the class conditional independence assumption is enforced.

[Ans: 14]

2024 Sep 01 ET

6. For a dataset with features in $\mathbb{R}^3$, which of the following expresses the class conditional independence assumption in a Naive Bayes model? $p(\cdot)$ denotes probability.

(a) $p((x_1, x_2, x_3) \mid y) = p(x_1 \mid y) \cdot p(x_2 \mid y) \cdot p(x_3 \mid y)$

(b) $p((x_1, x_2, x_3), y) = p(y) \cdot p(x_1) \cdot p(x_2) \cdot p(x_3)$

(c) $p((x_1, x_2, x_3) \mid y) = \frac{p((x_1,x_2,x_3),y)}{p(y)}$

(d) $p(y \mid (x_1, x_2, x_3)) = \frac{p((x_1,x_2,x_3),y)}{p(x_1,x_2,x_3)}$

[Ans: a]

2024 Dec 01 Quiz 2

7. **Comprehension:**
Based on the above data, answer the given subquestions.
A binary classification dataset has 2000 data points belonging to $\{0,1\}^2$. A Naive Bayes algorithm was run on the same dataset, resulting in the following estimates:

$$\begin{aligned}
\hat{p}, &\text{ estimate for } P(y=1) = 0.4 \\
\hat{p}_1^0, &\text{ estimate for } P(f_1 = 1 \mid y = 0) = 0.25 \\
\hat{p}_2^0, &\text{ estimate for } P(f_2 = 1 \mid y = 0) = 0.35 \\
\hat{p}_1^1, &\text{ estimate for } P(f_1 = 1 \mid y = 1) = 0.15 \\
\hat{p}_2^1, &\text{ estimate for } P(f_2 = 1 \mid y = 1) = 0.05
\end{aligned}$$

i. What is the estimated value of $P(f_2 = 0 \mid y = 1)$? Write your answer correct to two decimal places.

[Ans: 0.94 to 0.96]

ii. What will be the predicted label for the data point $[1, 0]$?

[Ans: 0]

8. Suppose you want to use a Naive Bayes classifier to predict whether a student will pass or fail an exam based on two features: the number of hours they studied and whether they attended review sessions. Assume that the features are conditionally independent given the exam outcome and that the variances of the study hours distributions are equal for both pass and fail categories.
How many parameters are required to classify a new student using this Naive Bayes classifier?

[Ans: 6]

2024 Dec 22 ET

9. **Comprehension:**
Consider the following dataset for a binary classification problem in which the features are from $\{0,1\}^3$ and the labels are from $\{0,1\}$.

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

A Naive Bayes classifier is trained on this dataset. The parameters to be estimated are represented as $\hat{p}_i^y$, which are presented in the form of the table given below:

i	$y=0$	$y=1$
1	a	b
2	c	d
3	e	f

Assume that there is no smoothing. Answer the given sub-questions.

i. Find the value of $a+b$.

[Ans: 1]

ii. Find the value of $b+d+f$.

[Ans: 2]

iii. Find the predicted label for the test point $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

[Ans: 1]

2025 Mar 16 Quiz 2

10. Consider a k-class classification problem with five binary features f_1, f_2, f_3, f_4, and f_5, where $k > 2$. How many parameters need to be estimated to learn a Naive Bayes classifier?

(a) $5k$

(b) $6k - 1$

(c) $(6k - 1)^5$

(d) $3k - 3$

[Ans: b]

11. What is the main purpose of using Laplace smoothing in Naïve Bayes?

(a) To prevent numerical errors when calculating the variance of Gaussian features.

(b) To ensure that unseen categorical feature values receive small but nonzero probabilities.

(c) To strengthen the model's predictions by increasing the weight of certain features.

(d) To balance the likelihood values of all features before computing posterior probabilities.

[Ans: b]

12. Consider a Naive Bayes problem with three features, f_1, f_2 and f_3 for a binary classification problem. We have a total of 6 training samples given in the table below:

Sample	f_1	f_2	f_3	y
x_1	0	1	1	0
x_2	1	0	1	0
x_3	0	0	1	0
x_4	0	0	0	1
x_5	1	0	1	1
x_6	0	1	0	1

For $x_{\text{test}} = [1 \quad 1 \quad 0]^T$, what output would Naive Bayes model predict assuming that no Laplacian smoothing has been done as part of the algorithm?

[Ans: 1]

2025 Apr 13 ET

13. The following dataset represents different weather conditions and the corresponding decision on whether a person chooses to play tennis:

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Overcast	Hot	High	Weak	No
2	Sunny	Hot	High	Weak	No
3	Sunny	Hot	Low	Weak	Yes
4	Overcast	Hot	High	Weak	Yes
5	Overcast	Cool	Low	Strong	Yes
6	Sunny	Cool	High	Weak	Yes
7	Sunny	Cool	Low	Weak	No
8	Sunny	Cool	Low	Strong	No
9	Sunny	Hot	Low	Strong	Yes
10	Overcast	Hot	High	Weak	Yes

Using the Naïve Bayes algorithm, compute the probability that a person will play tennis (Yes) given the conditions:

- Outlook = Overcast
- Temperature = Cool
- Humidity = High
- Wind = Weak

Enter the answer up to three decimal places.

[Ans: 0.61 to 0.67]

2025 Aug 03 Quiz 2

14. Which among the following are true? (MSQ)

(a) In generative model, we are interested in how the features are generated.

(b) In discriminative model, we are interested in how the features are generated.

(c) In discriminative model, we are interested in modelling $p(y \mid x)$, where y is the label and x is a feature.

(d) In generative model, we are interested in modelling $p(y \mid x)$, where y is the label and x is a feature.

(e) Gaussian Naive Bayes is a discriminative model.

[Ans: a, c]

15. **Comprehension:**
Based on the above data, answer the given subquestions.
Consider a 2-class Naive Bayes classifier with first feature f_1 as binary and the second feature f_2 as Gaussian. Following are the information that are provided to us:

$$\begin{aligned} \text{Estimate of } P(Y=1), \hat{p} &= 0.4 \\ \text{Estimate of } P(f_1=1 \mid y=0), \hat{p}_1^0 &= 0.4 \\ \text{Estimate of } P(f_1=1 \mid y=1), \hat{p}_1^1 &= 0.5 \\ (f_2 \mid y=0) &\sim \mathcal{N}(0,1) \\ (f_2 \mid y=1) &\sim \mathcal{N}(1,1) \end{aligned}$$

i. If a test point has label 0, what will be the probability that the point is $x_{\text{test}} = [0 \quad 0]^T$? Enter the answer correct to two decimal places.

[Ans: 0.22 to 0.26]

ii. For $x_{\text{test}} = [0 \quad 1.5]^T$, what output would Naive Bayes model predict?

[Ans: 1]

2025 Aug 31 ET

16. **Comprehension:**
Consider a naive Bayes classification problem with three binary features f_1, f_2, f_3 for a binary classification problem. We have a total of 6 training samples given in the table below:

Sample	f_1	f_2	f_3	y
x_1	1	1	0	0
x_2	1	0	1	0
x_3	1	0	0	0
x_4	0	0	1	1
x_5	0	0	0	1
x_6	0	1	0	1

Based on the above data, answer the given subquestions.

i. How many parameters need to be estimated to make a prediction for a data point using a naive Bayes algorithm?

[Ans: 7]

ii. If a test point has label 1, what is the probability of $x_{\text{test}} = [1 \quad 1 \quad 1]^T$? Assume that Laplacian smoothing is done, i.e., add two pseudo training examples $x^0_{\text{pseudo}} = [1 \quad 1 \quad 1]^T$ for class 0 and $x^1_{\text{pseudo}} = [1 \quad 1 \quad 1]^T$ for class 1. Enter the answer correct to four decimal places.

[Ans: 0.0625]

2025 Dec 18 ET

17. Consider a naive Bayes classification problem with three binary features f_1, f_2, f_3 for a binary classification problem. You are given the following 8 training samples:

Sample	f_1	f_2	f_3	y
x_1	0	1	0	0
x_2	1	1	0	0
x_3	1	1	1	1
x_4	1	0	0	1
x_5	1	0	1	0
x_6	0	1	1	1
x_7	1	1	0	1
x_8	0	0	1	0

Compute the Naive Bayes parameter estimate $\hat{p}(f_2 = 1 \mid y = 1)$ (without Laplace smoothing). Enter your answer correct to two decimal places.

[Ans: 0.75]

2026 Apr 06 Quiz 2

18. **Comprehension:**
You are provided with a training set of 5 observations, each consisting of four binary features $X = \{X_1, X_2, X_3, X_4\}$ and a binary target label $Y \in \{0, 1\}$.

Sample ID	X_1	X_2	X_3	X_4	Y
1	1	0	1	1	1
2	0	1	1	0	0
3	1	1	0	1	1
4	0	0	1	1	1
5	1	0	0	0	0

Based on the above data, answer the given subquestions.

i. Using the Naive Bayes assumption and the data provided, calculate the joint probability $P(X, Y = 1)$ for a new test instance $X = (1, 0, 1, 1)$. Enter your answer correct to two decimal places.

[Ans: 0.15 to 0.21]

ii. To address zero-probability estimates using Laplace smoothing, the dataset is augmented with two synthetic observations: one sample with all features $(X_1, X_2, X_3, X_4) = 1$ and $Y = 1$, and one sample with all features $(X_1, X_2, X_3, X_4) = 1$ and $Y = 0$. Based on the new augmented dataset of 7 samples, calculate the updated conditional probability $P(X_2 = 1 \mid Y = 1)$.

[Ans: 0.5]

iii. For the Naive Bayes model with n binary features and a binary class label, how many independent parameters are required to fully specify the model?

(a) 5
(b) 8
(c) 9
(d) 10

[Ans: c]

2026 Aug 14 Quiz 2

19. Suppose an email is represented by a binary feature vector $X = (x_1, x_2, \ldots, x_d) \in \{0, 1\}^d$. A generative classifier first chooses the class label Y according to the prior distribution $P(Y)$. It then generates the feature vector X conditioned on the chosen class. The classifier assumes that, given the class label, the features are conditionally independent. Which of the following expressions correctly represents $P(Y = \text{Spam} \mid X = x)$?

(a) $\dfrac{P(Y = \text{Spam})}{P(X = x)}$

(b) $\dfrac{\left(\prod_{i=1}^{d} P(x_i)\right) P(Y = \text{Spam})}{P(X = x)}$

(c) $\prod_{i=1}^{d} P(Y = \text{Spam} \mid x_i)$

(d) $\dfrac{\left(\prod_{i=1}^{d} P(x_i \mid Y = \text{Spam})\right) P(Y = \text{Spam})}{P(X = x)}$

[Ans: d]

20. A generative classifier models the joint distribution of the feature vector X and the class label Y. For a particular email represented by the feature vector x, suppose $P(X = x, Y = \text{Spam}) = 0.015$ and $P(X = x, Y = \text{Not Spam}) = 0.035$. Compute the posterior probability $P(Y = \text{Spam} \mid X = x)$.

[Ans: 0.3]

2026 Sep 13 ET

21. Given a dataset with K discrete attributes, each taking exactly 5 possible values (where $K > 2$), for a three-class classification task, the number of parameters to be estimated for learning a naïve Bayes classifier is:

(a) $3K + 2$

(b) $12K + 2$

(c) $3 \cdot 2^K$

(d) K^3

[Ans: b]