Manoj Kumar

GATE AIR - 13

M.Tech in Data Science From IIT Guwahati

Expertise in Machine Learning, Deep Learning, Artificial Intelligence, Probability and Statistics

Telegram: @Manoj_Gate_DSAI

Unacademy

CLICK

Some True/False ?

1 If events A and B are conditionally independent given C, then A and B are also independent of each other. (True/False)?

2. If A and {B, C} are conditionally independent given D, are A and B conditionally independent given D? (True/False)?

3. Maximizing the likelihood of a logistic regression model yields multiple local optima. (True/False)?

4. For a fixed size of the training and test set, increasing the complexity of the model always leads to a reduction of the test error. (True/False)?

5. Linear regression gives the Maximum Likelihood Estimator (MLE) for data from a model where y_i is normally distributed as $N\left(\sum_j w_j x_{ij}, \sigma^2\right)$. (True/False)?

Some True/False ?

6 Logistic Regression uses the logistic function $\sigma(z) = \frac{1}{1+e^{-z}}$, where z is a linear combination of the input features. Will logistic regression always give a linear decision boundary? (True/False)?

7. In machine learning, the bias is always a bigger source of error than the variance. (True/False)?

8. Suppose a dataset is linearly separable. Is a logistic regressor with regularization parameter $\lambda > 0$ guaranteed to separate the data? (True/False)?

9. In ordinary least squares regression, the coefficient estimator $\hat{w} = (X^T X)^{-1} X^T Y$ that minimizes the residual sum of squares is also the maximum likelihood estimator under the assumption that the errors are normally distributed with constant variance. (True/False)?

Question 10

In a classification problem with two classes C_1 and C_2, the posterior probability for class C_1 is defined as:

$$P(C_1|X) = \frac{1}{1+\exp(-a)} = \sigma(a)$$

What is the value of a?

Options:

1. $a = \ln\left(\frac{P(X|C_1)P(C_1)}{P(X|C_2)P(C_2)}\right)$
2. $a = \ln\left(\frac{P(X|C_2)P(C_2)}{P(X|C_1)P(C_1)}\right)$
3. $a = \ln\left(\frac{P(C_1|X)}{P(X|C_2)P(C_2)}\right)$
4. $a = \ln\left(\frac{P(X|C_2)}{P(X|C_1)}\right)$

https://t.me/gatedsai2025

Question 11

In logistic regression, if the predicted logit value is 0, what is the corresponding transformed probability?

(A) 0
(B) 1
(C) 0.5
(D) 0.05

Question 12

Consider a naive Bayes classifier with three Boolean input variables, X_1, X_2, and X_3, and one Boolean output variable, Y. How many parameters must be estimated to train this classifier?

Question 13

Discriminative models are used to estimate which of the following probabilities? Assume x denotes the input features and y the output labels.

(a) $P(y \mid x)$
(b) $P(y, x)$
(c) $P(x \mid y)$
(d) $P(y)$

Question 14

The solution to the following L_1 regularized least-squares regression problem

$$\arg\min_{\mathbf{w}} \|\mathbf{Y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1$$

for $\lambda > 0$ is:

(a) $(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$

(b) $(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{Y}$

(c) The objective is unbounded, i.e., the solution is $-\infty$.

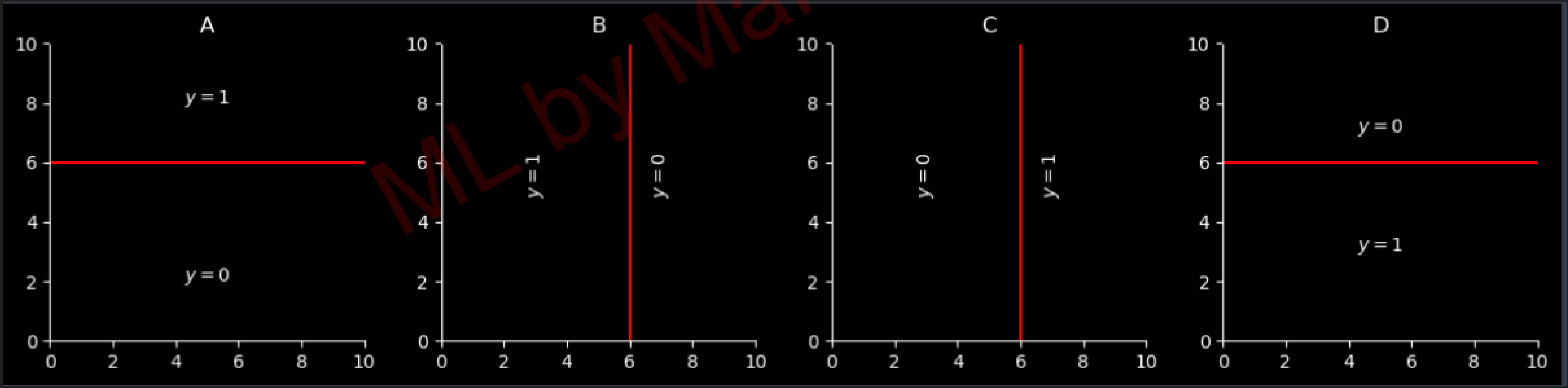
(d) None of the above

Question 15

Suppose you train a logistic regression classifier and the learned hypothesis function is

$$h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2),$$

where $\theta_0 = 6$, $\theta_1 = 0$, $\theta_2 = -1$. Which of the following represents the decision boundary for $h_\theta(x)$?



Question 16

A machine learning model exhibits poor performance on both the training and test datasets. What might explain this outcome?

A) The model is intricately fitting nuances in the training data, leading to high variance.

B) The model lacks sufficient complexity to capture the fundamental patterns in the data.

C) The model performs perfectly on the training data but fails on the test data.

D) The model's complexity is excessively high relative to the volume of available training data.

Question 17

Which of the following is true about the Naive Bayes classifier?

A) The Naive Bayes classifier requires that features must be identically distributed within each class to ensure accurate predictions.

B) The classifier assumes that features are independent across different classes.

C) The Naive Bayes classifier assumes that all input features are conditionally independent given the class.

D) The classifier can only be applied to datasets where input features are continuously and normally distributed.

Question 18

Suppose you are given the following set of data with three Boolean input variables a, b, and c, and a single Boolean output variable K.

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	0	0
0	0	1	1
0	0	1	1
0	0	0	0

Assume we are using a naive Bayes classifier to predict the value of K from the values of the other variables.

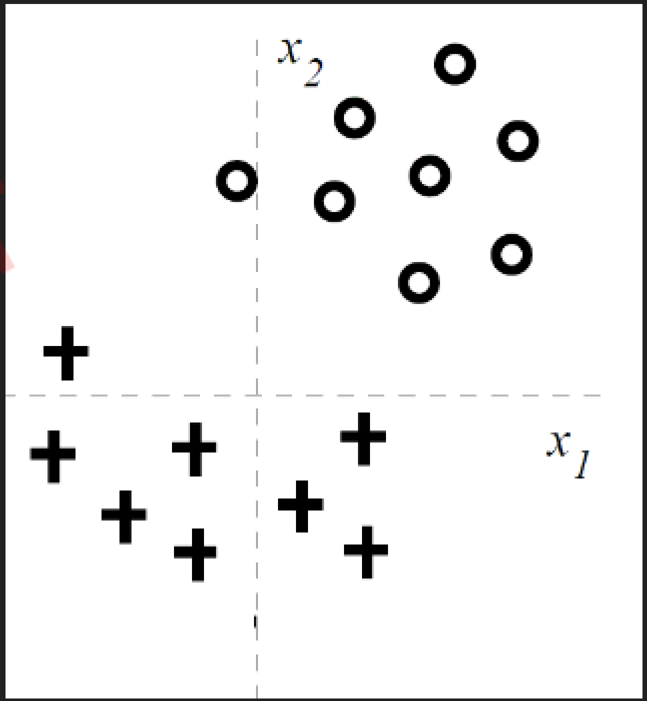
what is $P(K = 1|a = 1, b = 1, c = 0)$?

Question 19[MSQ]

https://t.me/gatedsai2025

Given the binary classification task depicted in the figure with the simple linear logistic regression model:

$$P(y=1|\vec{x},\vec{w}) = g(w_0 + w_1x_1 + w_2x_2) = \frac{1}{1+\exp(-w_0-w_1x_1-w_2x_2)}$$



Notice that the training data can be separated with zero training error using a linear separator. Consider training regularized linear logistic regression models where we try to maximize the following objective function:

$$\sum_{i=1}^{n} \log\left(P(y_i|x_i, w_0, w_1, w_2)\right) - Cw_j^2$$

where C is very large and only one of the parameters w_j is regularized in each case.

How does the number of misclassifications change with regularization of each parameter w_j, when C tends to infinity?

(a) Regularizing w_0 will decrease the number of misclassifications.

(b) Regularizing w_1 will increase the number of misclassifications.

(c) Regularizing w_2 will have no effect on the number of misclassifications.

(d) Regularizing w_0 will increase the number of misclassifications.

Question 20

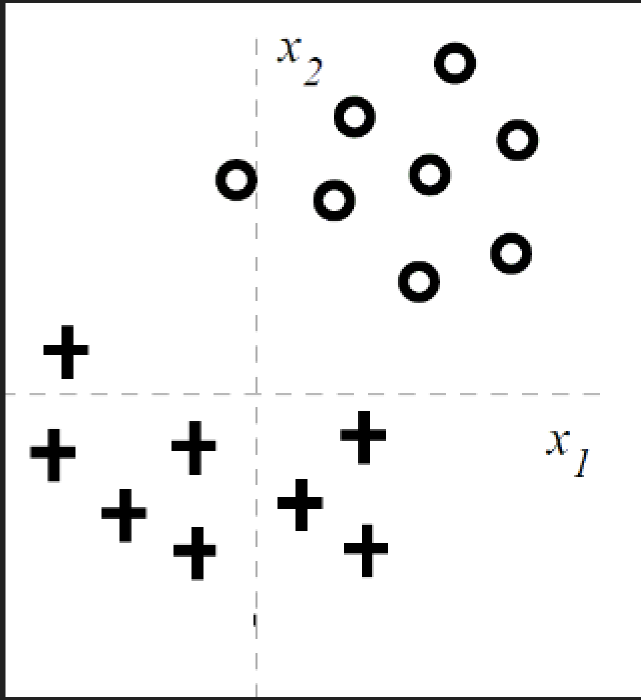
https://t.me/gatedsai2025

Consider the binary classification task depicted in the figure with the simple linear logistic regression model:

$$P(y=1|\vec{x},\vec{w}) = g(w_0 + w_1 x_1 + w_2 x_2) = \frac{1}{1+\exp(-w_0 - w_1 x_1 - w_2 x_2)}$$

The training data can be separated with zero training error using a linear separator. Consider training regularized linear logistic regression models where we try to maximize the following objective function:

$$\sum_{i=1}^{n} \log\left(P(y_i|x_i, w_0, w_1, w_2)\right) - C(|w_1| + |w_2|)$$



As we increase the regularization parameter C, which of the following statements can be true? (Choose only one):

(a) First w_1 will become 0, then w_2.

(b) First w_2 will become 0, then w_1.

(c) w_1 and w_2 will become zero simultaneously.

(d) None of the weights will become exactly zero, only smaller as C increases.

Question 21

Consider a regularized linear logistic regression model given by:

$$P(y = 1|\mathbf{x}, \mathbf{w}) = g(w_0 + w_1x_1 + w_2x_2)$$

We aim to maximize the following regularized log-likelihood:

$$\sum_{i=1}^{n} \log p(y_i|\mathbf{x}_i, w_0, w_1, w_2) - Cw_j^2$$

where j can be 0, 1, or 2.

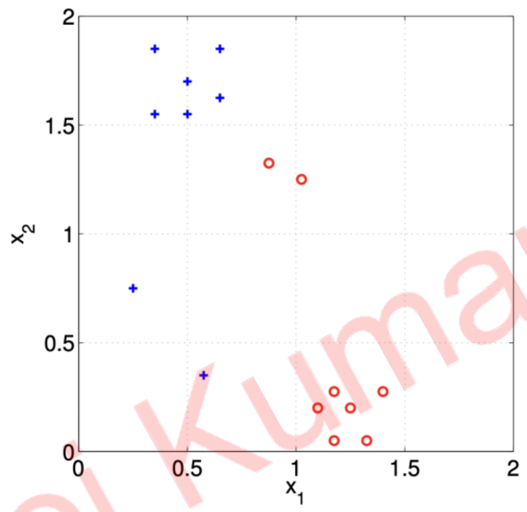


Figure 4.1 Labeled training set (reproduced here for clarity)

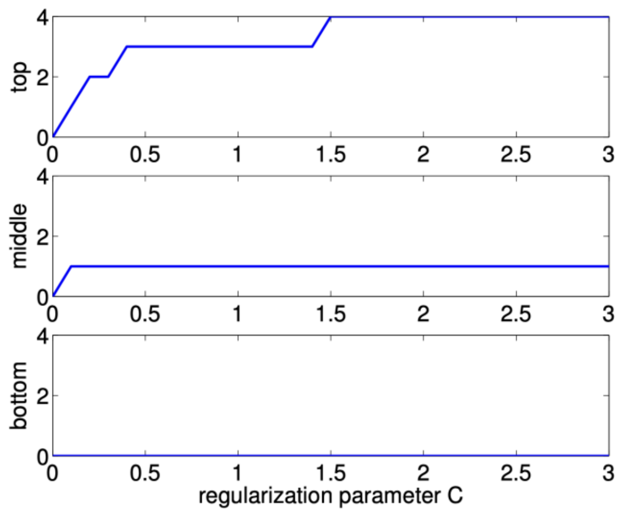


Figure 4.2. Training errors as a function of regularization penalty

Based on the labeled training set in Figure 4.1 and the resulting training errors (number of misclassifications) as a function of the regularization parameter C in Figure 4.2, assign the "top," "middle," and "bottom" plots to the correct parameter w_0, w_1, or w_2 that was regularized in the plot.

A. Top: w_0, Middle: w_1, Bottom: w_2

B. Top: w_1, Middle: w_2, Bottom: w_0

C. Top: w_2, Middle: w_1, Bottom: w_0

D. Top: w_1, Middle: w_0, Bottom: w_2

Question 22

Consider the following logistic regression models for binary classification with a sigmoid function $g(z) = \frac{1}{1+e^{-z}}$:

- **Model 1:** $P(Y = 1|X, w_1, w_2) = g(w_1 X_1 + w_2 X_2)$
- **Model 2:** $P(Y = 1|X, w_1, w_2) = g(w_0 + w_1 X_1 + w_2 X_2)$

Given three training examples:

- $x^{(1)} = [1, 1]^T, y^{(1)} = 1$
- $x^{(2)} = [1, 0]^T, y^{(2)} = -1$
- $x^{(3)} = [0, 0]^T$, originally $y^{(3)} = 1$

If the label of the third example is changed to -1, does it affect the learned weights $w = (w_1, w_2)$ in Model 1 and Model 2?

A. It affects both Model 1 and Model 2.

B. It affects neither Model 1 nor Model 2.

C. It affects only Model 1.

D. It affects only Model 2.

Question 23

Consider a logistic regression model where the hypothesis $h_\theta(x)$ is given by:

$h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$

where $\sigma(z)$ is the sigmoid function defined as $\sigma(z) = \frac{1}{1+e^{-z}}$.

The model includes quadratic transformations of the features x_1^2 and x_2^2, and a cross-term $x_1 x_2$. The coefficients for these features are set as follows: $\theta_3 = 2$, $\theta_4 = 2$, and $\theta_5 = 0$. Given these settings, what shape will the decision boundary most likely resemble?

A) A straight line

B) A circle or ellipse

C) A parabola

D) A hyperbola

Question 24

Consider a Gaussian Naive Bayes classifier for a dataset with a single attribute x and two classes 0 and 1. We classify a point $\mathbf{x}$ as class 1 if:

$P(y=1|x) \geq P(y=0|x)$

For this Gaussian Naive Bayes classifier, the parameters for the two Gaussian distributions are:

$x|y=0 \sim N(0, 1/4)$

$x|y=1 \sim N(0, 1)$

$P(y=1) = 0.5$

Here, $N(\mu, \sigma^2)$ represents a normal distribution with mean μ and variance σ^2.

Determine the decision boundary for classifying a point $\mathbf{x}$ as class 1.

(a) $x^2 \leq \frac{2\ln 2}{3}$

(b) $x \leq \sqrt{\frac{2\ln 2}{3}}$

(c) $x^2 > \frac{2\ln 2}{3}$

(d) $x \geq \sqrt{\frac{2\ln 2}{3}}$

https://t.me/gatedsai2025

Question 25

https://t.me/gatedsai2025

Consider the binary classification problem where class label $Y \in \{0,1\}$ and each training example X has 2 binary attributes $X_1, X_2 \in \{0,1\}$. The class priors are $P(Y=0) = P(Y=1) = 0.5$, and the conditional probabilities are given as follows:

$P(X_1|Y)$

	$X_1 = 0$	$X_1 = 1$
$Y = 0$	0.7	0.3
$Y = 1$	0.2	0.8

$P(X_2|Y)$

	$X_2 = 0$	$X_2 = 1$
$Y = 0$	0.9	0.1
$Y = 1$	0.5	0.5

The expected error rate is the probability that a classifier provides an incorrect prediction for an observation. If Y is the true label, let $\hat{Y}(X_1, X_2)$ be the predicted class label. The expected error rate is:

$$P_D\left(Y \neq \hat{Y}(X_1, X_2)\right) = \sum_{X_1=0}^{1} \sum_{X_2=0}^{1} P_D\left(X_1, X_2, Y = 1 - \hat{Y}(X_1, X_2)\right)$$

Compute the expected error rate of this naive Bayes classifier which predicts Y given both of the attributes X_1, X_2. Assume that the classifier is learned with infinite training data.

Question 26

You are given a dataset with N records in which the i-th record has a real-valued input attribute x_i and a real-valued output y_i, which is generated from a Gaussian distribution with mean $\sin(wx_i)$ and variance 1.

$$P(y_i|w,x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_i - \sin(wx_i))^2}{2}\right)$$

We have one unknown parameter w in the above model and we want to learn the maximum likelihood estimate of it from the data.

If you compute the maximum likelihood estimate of the parameter w, which of the following equations is satisfied?

(a) $\sum_i \cos^2(wx_i) = \sum_i y_i \sin(x_i)$

(b) $\sum_i \cos^2(wx_i) = \sum_i y_i \sin(2wx_i)$

(c) $\sum_i x_i \sin(wx_i) \cos(wx_i) = \sum_i y_i \cos(wx_i)$

(d) $\sum_i x_i \cos(x_i) = \sum_i y_i \cos(wx_i)$

Question 27

Let $F(x) = w_0 + \sum_{j=1}^{d} w_j x_j$ and $L(y^i F(x^i)) = \frac{1}{1+\exp(y^i F(x^i))}$ be the cost function to be minimized. Here, d represents the number of features, x_j represents the j-th feature of the input x, y is the target variable, and i represents the index of the training sample.

Suppose you use gradient descent to obtain the optimal parameters w_0 and w_j. Give the update rules for these parameters.

Which of the following is the correct update rule for the parameter w_k?

(a) $w_k^{(t+1)} = w_k^t - \eta \sum_i x_k^i \left(\frac{\exp(y^i F(x^i))}{(1+\exp(y^i F(x^i)))^2} \right)$

(b) $w_k^{(t+1)} = w_k^t + \eta \sum_i y^i x_k^i \left(\frac{\exp(y^i F(x^i))}{(1+\exp(y^i F(x^i)))^2} \right)$

(c) $w_k^{(t+1)} = w_k^t - \eta \sum_i y^i x_k^i \left(\frac{\exp(y^i F(x^i))}{(1+\exp(y^i F(x^i)))^2} \right)$

(d) $w_k^{(t+1)} = w_k^t + \eta \sum_i x_k^i \left(\frac{\exp(y^i F(x^i))}{(1+\exp(y^i F(x^i)))^2} \right)$