Machine Learning

Linear Classification & Logistic Regression

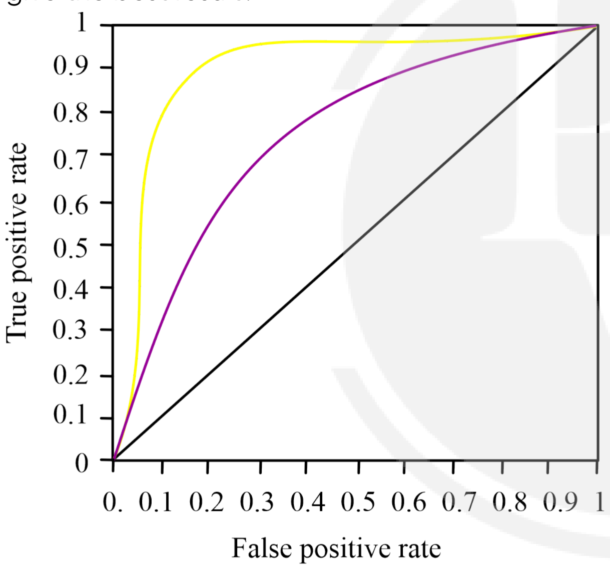
DPP : 1

Q1 In a logistic regression(Linear Classifier) problem, what is a possible output for a new instance?

(A) 85 (B) -0.19

(C) 1.20 (D) 89%

Q2 The below figure shows AUC-ROC curves for three logistic regression models. Different colors show curves for different hyper parameters values. Which of the following AUC-ROC will give the best result?



(A) Yellow (B) Pink

(C) Black (D) All are same

Q3 In the regression model (y = a + bx) where x = 2.50, y = 5.50 and a = 1.50 ($\bar{x}$ and $\bar{y}$ denote mean of variables x and y and a is a constant), which one of the following values of parameter 'b' of the model is correct?

(A) 1.75 (B) 1.60

(C) 2.00 (D) 2.50

Q4 Which of the following is an advantage of linear classification algorithms?

(A) They are highly interpretable

(B) They can capture complex non-linear relationships in the data

(C) They are less sensitive to outliers compared to other algorithms

(D) They require less computational resources for training and prediction

Q5 The learner is trying to predict housing prices based on the size of each house. What type of regression is this?

(A) Multivariate Logistic Regression

(B) Logistic Regression

(C) Linear Regression

(D) Multivariate Linear Regression

Q6 The hypothesis is given by h(x) = t0 + t1x. What is the goal of t_0 and 0_1?

(A) Give negative h(x)

(B) Give h(x) as close to 0 as possible, without themselves being 0

(C) Give h(x) as close to y, in training data, as possible

(D) Give h(x) closer to x than y

Q7 In continuation with question 7, let x = 1 if the server is wearing black shirt and x = 0 for servers wearing other colored shirts. We know that there are 2 points 70 observations with x 1 and 340 observations with x = 0. The response variable is also an indicator variable given by y = 1 if the customer left a tip and y=0 if the customer did not leave a tip. Use this data to fit a logistic regression model to compute the log-odds of leaving a tip depending on the color of the server's shirt.

(A) -0.4797 +0.1249x

(B) 0.2877 +0.1249x

(C) 0.1249+0.4317x

(D) -0.4797 +0.7674x

Q8 In Simple Logistic regression the predictor . . . ?

(A) is interval/ratio data

(B) must undergo a logarithmic transformation before undergoing logistic regression

(C) be in the range of 0 to 1

(D) represent ranked scores

(E) be a binary variable

Q9 In logistic regression the logit is . . . : (one correct choice)

(A) the natural logarithm of the odds ratio.

(B) an instruction to record the data.

(C) a logarithm of a digit.

(D) the cube root of the sample size.

Q10 Given an example from a dataset (x1, x2) = (4, 1), observed value y = 2 and the initial weights w1, w2, bias b as -0.015, -0.038 and 0. What will be the prediction y'.

(A) 0.01 (B) 0.03

(C) 0.05 (D) 0.1

Q11 For the linear regression model y_i- = β_0 + $\beta_1 x_i$, + let $\hat{y}_i$ denote the sample fitted value for the i-th unit from the OLS procedure based on a sample of n observations and $\widehat{\beta}_0$ and $\widehat{\beta}_1$ be the OLS estimators of β_0 and β_1, respectively. Which of the following is true?

(A) $\sum_{i=1}^{n} u_i = 0$

(B) $\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$.

(C) $E(\hat{y}|x) = \widehat{\beta}_0 + \widehat{\beta}_1 x$

(D) $E(y|x) = \widehat{\beta}_0 + \widehat{\beta}_1 x$

Q12 A classification table: (one correct choice)

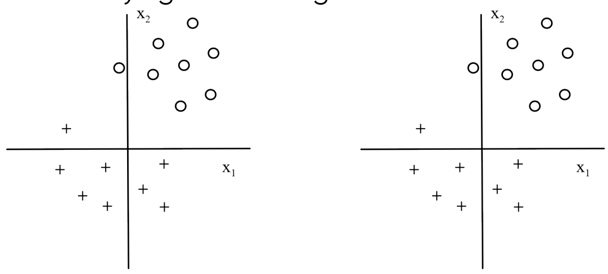
(A) helps the researcher assess statistical significance.

(B) indicates how well a model has predicted group membership.

(C) indicates how well the independent variable(s) correlate with the dependent variable.

(D) provides a basis for calculating the exp(b) value

Q13 Suppose you are given the below data, and you want to apply a logistic regression model for classifying it into two given classes.



You are using logistic regression with L1 regularization.

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

Where C is the regularization parameter, and w1 & w2 are the coefficients of x1 and x2.

Which of the following option is correct when you increase the value of C from zero to a very large value?

(A) First, w2 becomes zero, and then w1 becomes zero

(B) First, w1 becomes zero, and then w2 becomes zero

(C) Both become zero at the same time

(D) Both cannot be zero even after a very large value of C

Q14 Likelihood (In the statistical sense) . . (one correct choice)

(A) Is the same as a p value

(B) Is the probability of observing a particular parameter value given a set of data

(C) attempts to find the parameter value which is the most likely given the observed data.

(D) minimizes the difference between the model and the data

Q15 A Maximum Likelihood Estimator (in the statistical sense) . . (one correct choice)

(A) Is the same as a p value

(B) Is the probability of observing a particular parameter value given a set of data

(C) attempts to find the parameter value which is the most likely given the observed data.

(D) Is the same as R Square

Q16 Why cost function which has been used for linear regression can't be used for logistic regression?

(A) Linear regression uses mean squared error as its cost function. If this is used for logistic regression, then it will be a non-convex function of its parameters. Gradient descent will converge into the global minimum only if the function is convex.

(B) Linear regression uses mean squared error as its cost function. If this is used for logistic regression, then it will be a convex function of its parameters. Gradient descent will converge into the global minimum only if the function is convex.

(C) Linear regression uses mean squared error as its cost function. If this is used for logistic regression, then it will be a non-convex function of its parameters. Gradient descent will converge into the global minimum only if the function is non-convex.

(D) Linear regression uses mean squared error as its cost function. If this is used for logistic regression, then it will be a convex function of its parameters. Gradient descent will converge into the global minimum only if the function is non-convex.

Answer Key

Q1	(A)
Q2	(A)
Q3	(B)
Q4	(A)
Q5	(C)
Q6	(C)
Q7	(D)
Q8	(E)
Q9	(A)
Q10	(D)
Q11	(A)
Q12	(B)
Q13	(B)
Q14	(C)
Q15	(C)
Q16	(A)

Hints & Solutions

Q1 Text Solution:

The output can only be between 0 and 1.

Q3 Text Solution:

$(y = a + bx)$

where,

- $\bar{x} = 2.50$
- $\bar{y} = 5.50$
- $a = 1.50$
- ($\bar{x}$ and $\bar{y}$ denote mean of variables x and y and a is a constant)

Putting values in the formula:

$5.50 = 1.50 + b \times 2.50$

$b \times 2.50 = 4$

$b = 4/2.5 = 1.60$

Q8 Text Solution:

Logistic regression is commonly used when the outcome or dependent variable is binary (e.g., yes/no, 0/1), and it models the probability of the outcome occurring as a function of the predictor variable.

Q10 Text Solution:

Given

$x^1 = 4$, $x^2 = 1$, $w1 = -0.015$, $w^2 = -0.038$, $y = 2$ and $b = 0$.

Then prediction $y' = w^1 x1 + w^2 x^2 + b$

$= (-0.015 * 4) + (-0.038 * 1) + 0$

$= -0.06 + -0.038 + 0$

$= -0.098$

$= -0.1$

Q11 Text Solution:

The equation states that the sum of the residuals (ui) in a linear regression model is equal to zero. This is a fundamental property of the Ordinary Least Squares (OLS) method used to estimate the coefficients in a linear regression model.

Q12 Text Solution:

A classification table, also known as a confusion matrix, shows how well a model has predicted group membership or class labels. It compares the actual class labels with the predicted class labels generated by the model.

Q13 Text Solution:

By looking at the image, we see that even by just using x2, we can efficiently perform classification. So at first, w1 will become 0. As the regularization parameter increases more, w2 will come closer and closer to 0.

Q14 Text Solution:

Likelihood, in the statistical sense, refers to the probability of observing a particular parameter value given a set of data. It represents how well the parameter value fits the observed data.

Q15 Text Solution:

A Maximum Likelihood Estimator (MLE) is a method used to estimate the parameters of a statistical model. It aims to find the parameter values that maximize the likelihood function, making them the most likely given the observed data.

Q16 Text Solution:

- In linear regression, the mean squared error (MSE) is used as the cost function.
- The MSE creates a convex optimization problem, meaning that it has a single minimum and gradient descent can converge to the global minimum efficiently.
- However, when the MSE is used for logistic regression, it creates a non-convex optimization problem due to the logistic function used in logistic regression. This non-convexity can cause gradient descent to converge to local minima rather than the global minimum.
- Therefore, using the MSE as the cost function for logistic regression is not appropriate, and alternative cost functions like cross-entropy loss are used instead.