Unit 6 - Week 3 - Linear, logisitic regression, and multilayer perceptron

Assignment 03

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2019-08-21, 23:59 IST.

Instructions:

1. **Attempt all questions.**
2. **Submission deadline:** 21st August 2019 23:59 IST
3. **Solutions to be posted:** 22nd August 2019
4. Older browsers might show **unnecessary vertical bars** at the end of math equations.

1) Consider a data set with five predictors, $X_1 =$ Percentage of marks (Out of 100), $X_2 =$ IQ, $X_3 =$ Gender (1 for female and 0 for male), $X_4 =$ Interaction between percentage of marks and IQ, and $X_5 =$ Interaction between percentage of marks and gender. The response of the system is the starting salary after graduation (in thousands of rupees). Let the least squares fit of the model be given by the parameters, $\hat{\beta}_0 = 150, \hat{\beta}_1 = 0.8, \hat{\beta}_2 = 0.07, \hat{\beta}_3 = 15, \hat{\beta}_4 = 0.01, \hat{\beta}_5 = -1.2$. For a fixed value of IQ and percentage, which of the following is true. **2 points**

- Males always earn more than females.
- Females always earn more than males.
- Females earn more than males provided that the percentage is high enough.
- Males earn more than females provided that the percentage is high enough.

No, the answer is incorrect.
Score: 0
Accepted Answers:
Males earn more than females provided that the percentage is high enough.

2) With reference to question 1, the predicted salary (converted into rupees) of a female with 80 percentage of marks and IQ of 115 is **1 point**

- Rs. 2,74,500
- Rs. 3,14,050
- Rs. 2,33,050
- Rs. 2,89,500

No, the answer is incorrect.
Score: 0
Accepted Answers:
Rs. 2,33,050

3) Choose which of the following statements are true. **2 points**

- If two variables are correlated, they always have a linear relationship.
- A neural network architecture can be used to design a logistic regression algorithm.
- Logistic regression is mainly used for regression.
- Linear regression is sensitive to outliers.

No, the answer is incorrect.
Score: 0
Accepted Answers:
A neural network architecture can be used to design a logistic regression algorithm.
Linear regression is sensitive to outliers.

4) (True/False) The odds of getting a tail in a fair coin is 0.5. **1 point**

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
False

5) Consider a simple linear regression model (1-variable case). If we change the input variable by one unit, the output variable changes by **1 point**

- One unit.
- The intercept.
- The slope.
- No change.

No, the answer is incorrect.
Score: 0
Accepted Answers:
The slope.

6) Suppose the true relationship between X and Y is given by $Y = \alpha_0 + \alpha_1 X + \epsilon$. The given data is fit with a linear regression model, as well as, a cubic regression ($Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \epsilon$) model. Let the residual sum of squares for the linear regression be LRSS and the residual sum of squares for the cubic regression be CRSS. Which of the following statements are true? **2 points**

- During training, LRSS will be lower than CRSS when $\epsilon \neq 0$.
- During training, LRSS will be equal to CRSS when $\epsilon = 0$.
- During testing, LRSS will be lower than CRSS when $\epsilon \neq 0$.
- During testing, LRSS will be lower than CRSS when $\epsilon = 0$.

No, the answer is incorrect.
Score: 0
Accepted Answers:
During training, LRSS will be equal to CRSS when $\epsilon = 0$.
During testing, LRSS will be lower than CRSS when $\epsilon \neq 0$.

7) Consider the data collected from 410 customers in a restaurant. It is observed that 40 of the 70 customers tipped the server who was wearing a black shirt and 130 of the 340 customers tipped the server who was wearing a different color. Compute the logit or log-odds of tipping a server wearing a black shirt. **2 points**

- 0.2877
- 0.1249
- -0.7677
- -1.7677

No, the answer is incorrect.
Score: 0
Accepted Answers:
0.2877

8) 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 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.. **2 points**

- $-0.4797 + 0.1249x$
- $0.2877 + 0.1249x$
- $0.1249 + 0.4317x$
- $-0.4797 + 0.7674x$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$-0.4797 + 0.7674x$

9) Consider a simple linear regression model with noise $y_i = \alpha_0 + \alpha_1 x_i + \epsilon_i$ and suppose that we redefine the regressor variable x_i as the deviation from its own average $\bar{x}$, say $x_i - \bar{x}$ then the new intercept of the linear regression model is. **2 points**

- $\alpha_0 + \alpha_1 \bar{x}$
- $\alpha_0 - \alpha_1 \bar{x}$
- $\alpha_1 + \alpha_0 \bar{x}$
- $\alpha_1 - \alpha_0 \bar{x}$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\alpha_0 - \alpha_1 \bar{x}$

10) Consider a simple linear regression model $y_i = \bar{x}_i^T \bar{\beta} + \epsilon_i$ where $\bar{x}_i = [1, x_{i1}, \ldots, x_{ik}]$, ϵ_i is a Gaussian random variable with zero mean and unit variance, and $\bar{\beta} = [\beta_1, \ldots, \beta_k]$.The response variable y_i takes only 0 and 1 and y_i is a Bernoulli random variable with probability distribution $P(y_i = 1) = \pi_i$ and $P(y_i = 0) = 1 - \pi_i$. Then, **2 points**

- $\bar{x}_i^T \bar{\beta} = 1 - \pi_i$
- $\bar{x}_i^T \bar{\beta} = \pi_i$
- $\bar{x}_i^T \bar{\beta} = \frac{\pi_i}{2}$
- $\bar{x}_i^T \bar{\beta} = \pi_i(1 - \pi_i)$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\bar{x}_i^T \bar{\beta} = \pi_i$

11) Consider the data given below. **1 point**

X	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
Y	0	0	0	0	0	0	1	0	1	0	0	0	1	1	1	1	1

X corresponds to the number of hours spent and Y corresponds to finishing the NPTEL assignment on time (binary values). If a list of scores against the number of hours spent was given instead of binary value of Y, then which technique would you use to fit the data?

- Linear regression
- Logistic regression
- Clustering
- Decision tree

No, the answer is incorrect.
Score: 0
Accepted Answers:
Linear regression

12) Consider the data given in the table in question 11. Fit a logistic regression model and compute the probability that a student will finish the NPTEL assignment in time if he works for 4.8 hours (You may use any coding language to find the coefficient values of the logistic model). **2 points**

No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Range) 0.48,0.58