Quiz

Due Nov 26 at 9:40am **Points** 30 **Questions** 22 **Available** Nov 26 at 9am - Nov 26 at 9:40am 40 minutes
Time Limit 30 Minutes

Instructions

Dear Participants,

This is an In-class quiz on Supervised Learning- Regression course. It has 22 questions and the duration is 30 mins.

Datasets to be used:

insurance-1.csv

Regards,

Program Office

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	16 minutes	2 out of 30

Score for this quiz: **2** out of 30
Submitted Nov 26 at 9:30am
This attempt took 16 minutes.

Question 1

1 / 3 pts

Insurance Dataset:

Select all the codes that generate dummy variables for the following variables in insurance dataset?

Column_names = ['sex', 'children', 'smoker', 'region']

Correct Answer

- [ ] df_encode = pd.get_dummies(data = df, prefix = None, columns = Column_names, drop_first= True, dtype='int')

- [ ] df_encode = pd.get_dummies(data = df, prefix = None, columns = Column_names, drop_first= True, dtype='str')

Correct!

- [x] df_encode = pd.get_dummies(data = df, columns = Column_names, drop_first=True, dtype='int8')

Correct Answer

- [ ] df_encode = pd.get_dummies(data = df, prefix = None, drop_first= True, columns = Column_names)

Unanswered

Question 2

0 / 3 pts

Using the beta estimates, predict the charges for the following customer profile:

Customer_id	Age	Sex	BMI	Children	Smoker	Region
2578	63	Male	45.9	2	Yes	northeast

- 10.1338
- 13.2456
- 12.918

Correct Answer

- 10.4614

Question 3

0 / 1 pts

How do we identify the presence of Multicollinearity?

- Using correlation matrix

You Answered

- Using variance inflation factor

Correct Answer

- Using correlation matrix and variance inflation factor
- None of these

Question 4

1 / 1 pts

If the slope and intercept of the regression line is 7 and 5 then the value of Y when x is 12 is:

- 67

Correct!

- 89

- 60
- 84

Unanswered

Question 5 0 / 1 pts

In the regression line $y = \beta_0 + \beta_1 x$, if β_0 is 0 then

Correct Answer

- The line of best fit passes through the origin
- Does not crosses the x axis
- The line of best fit cuts the X axis to the right of the Y axis
- The line of best fit cuts the X axis to the left of the Y axis

Unanswered

Question 6 0 / 1 pts

In a linear regression problem, if we add a feature in linear regression model and retrain the same model then

- If R Squared increases, this variable is significant.
- If R Squared decreases, this variable is not significant.

Correct Answer

- Individually R squared cannot tell about variable importance.
- None of these

Unanswered

Question 7 0 / 1 pts

Which of the following method(s) is used to predict the continuous dependent variable.

Correct Answer

- [ ] Simple Linear Regression

Correct Answer

- [ ] Multiple Linear Regression
- [ ] Logistic Regression
- [ ] Multinomial Regression

Unanswered

Question 8

0 / 1 pts

Logistic Regression transforms the output probability to be in a range of [0, 1]. Which of the following function is used by logistic regression to convert the probability in the range between [0,1].

Correct Answer

- [x] Sigmoid
- [ ] Mode
- [ ] Probit
- [ ] Square

Unanswered

Question 9

0 / 3 pts

[Insurance Dataset]: In the given insurance dataset, if a multiple linear regression model is built as below, what is the coefficient of sex_male?

charges ≈ age + bmi + children + sex_male + smoker_yes + region_northwest + region_southeast + region_southwest

Note: Convert sex, smoker and region into dummy variables for the levels given above.

Correct Answer

- ( ) 256.86
- (x) -131.3
- ( ) 475.5
- ( ) -352.96

Unanswered

Question 10

0 / 1 pts

In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. What will be the change in output?

- ( ) By 1
- ( ) No change
- ( ) By intercept

Correct Answer

By slope

Unanswered

Question 11 0 / 1 pts

Suppose I applied a logistic regression model on data and got training accuracy X and testing accuracy Y. Now I want to add few new features in data. Select option(s) which are correct in such case.

a. Training accuracy always decreases.

b. Training accuracy always increases or remain same.

c. Testing accuracy always decreases

d. Testing accuracy always increases or remain same

Correct Answer

Only b

Both b and d

Only d

Both a and b

Unanswered

Question 12 0 / 1 pts

Which of the following methods do we use to best fit the data in Logistic Regression?

Least square error

Correct Answer

Maximum Likelihood

Jaccard distance

Lease square error and Maximum likelihood

Unanswered

Question 13 0 / 1 pts

What is the range of logit function?

Correct Answer

(-∞, ∞)

(0, 1)

(-1, 1)

(0, ∞)

Unanswered

Question 14

0 / 3 pts

[Insurance Dataset]: If a multiple linear regression is built for the following variables, what will be the standard error for southeast region?

charges ≈ age + bmi + children + sex_male + smoker_yes + region_southeast + region_southwest

Correct Answer

415.338

478.692

423.888

419.678

Unanswered

Question 15

0 / 1 pts

What is the True positive rate for the given confusion matrix, represented in the form of table?

743	38
25	210

Correct Answer

89.36%

95.13%

10.63 %

28.26%

Unanswered

Question 16

0 / 1 pts

Logistic regression assumes a:

Linear relationship between continuous predictor variables and the outcome variable.

Correct Answer

Linear relationship between continuous predictor variables and the logit of the outcome variable.

Linear relationship between continuous predictor variables.

Linear relationship between observations.

Unanswered

Question 17 0 / 1 pts

A residual is defined as

Correct Answer

Y-Yˆ

Error sum of square

Regression sum of squares

Type 1 error

Unanswered

Question 18 0 / 1 pts

If 'Time' is used as the independent variable in a simple linear regression analysis, then which of the following assumption could be violated

There is a linear relationship between the independent and dependent variables

The residual variation is the same for all fitted values of Y

The residuals are normally distributed

Correct Answer

Successive observations of the dependent variable are uncorrelated.

Unanswered

Question 19 0 / 1 pts

Which of the following selection begins with the null mode?

Correct Answer

Forward Selection

- Backward Selection
- Mixed Selection
- None of these

Unanswered

Question 20 0 / 1 pts

Which of the following selection method cannot be used if number of variables is less than number of observations?

- Forward Selection
- Backward Selection
- Mixed Selection

Correct Answer

- Any of these can be used

Unanswered

Question 21 0 / 1 pts

What will be the output of the logistic regression?

Correct Answer

- Probability
- Scores of the regressors
- Log odds
- Probability and log odds

Unanswered

Question 22 0 / 1 pts

Which of the following is/are assumptions of Logistic regression?

- Linear relationship between dependent and independent variables
- Residuals are normally distributed

Correct Answer

- Little or no multicollinearity among the independent variables

Correct Answer

- [ ] Outcome or response variable can be binary or multinomial

Quiz Score: **2** out of 30