☺ Math506-Exam-II-KFUPM-Dr.Jamal ☺

Name/ID:

Q1: A way of predicting an output variable from one or more input variables is

A: correlation	C: regression
B: PCA	D: MSE

Q2: A visualization Python library with advanced plot features is

A: matplotlib	C: seaborn
B: Graphix	D: Gagagraphix

Q3: In linear regression models, an iterative algorithm that is used to minimize the least squares error is

A: Normal Equation	C: Intercept
B: PCA	D: Gradient Descent

Q4: The linear regression coefficients ($\boldsymbol{W}$) can be obtained by the solution of the normal equation, $\boldsymbol{W} =$

A: $(X^T X)^{-1} X^T Y$	C: $X^{-1} Y$
B: $X^T Y$	D: $(X^T X)^{-1} \boldsymbol{W}$

Q5: Which of the following transforms a categorical attribute ‘Education’ using one hot encoding?

A: df['Education'].replace({1:'A',2:'B',3:'C'})	C: df['Education'].one_hot({1:'A',2:'B',3:'C'})
B: pd.get_dummies(df,columns=['Education'])	D: pd.one_hot(df,columns=['Education'])

Q6: A logistic regression model is trained using

A: model.fit(X_test, y_test)	C: model.fit(X_train, y_train)
B: model.train(X_train, y_train)	D: model.solve(X_train, y_train)

Q7: Ridge and Lasso Models are imported using

A: sklearn.penalized_model	C: sklearn.linear_model
B: sklearn.regulazor_model	D: sklearn.reduced_model

Q8: In the code: regr = Ridge(alpha=450), the value of alpha sets the

A: regularization coefficient	C: intercept parameter
B: maximum number of iterations	D: number of sample records

Q9: Which code that is used for building training and testing sets is

A: training_testing_builder (X, y, test_size=0.30)	C: train_test_split (X, y, test_size=0.30)
B: training_testing_spliter (X, y, test_size=0.30)	D: training_testing_spliting (X, y, test_size=0.30)

Q10: MSE or SSE can be used in

A: Lasso Model	C: Linear Regression Model
B: Ridge Model	D: All of the Above

Q11: Classification problems are a type of

A: Supervised Learning	C: PCA Models
B: Unsupervised Learning	D: All of the Above

Q12: Descriptive analysis uses

A: multivariate plots	C: correlations
B: regressions	D: gradient descent iterations

Q13: The correlation of two attributes determines

A: which one should be the target variable	C: the intercept
B: the right choice of the z-score	D: the strength of their linear relationship

Q14: Which of the following is the strongest correlation score?

A: 0.1	C: - 0.85
B: 0.80	D: 0.83

Q15: Which of the following sentences is False about the principal component analysis?

A: It is a supervised learning method.	C: It projects the input data into a lower dimensional. linear space
B: It is used to reduce the number of attributes.	D: It is used for data visualization and exploration.

Q16: Which of the following is False about linear regression?

A: It is used for prediction.	C: It defines a relationship between independent and dependent variables.
B: It is unsupervised learning model.	D: It defines the dependent variable as a linear function of independent variables.

Q17: In a single input, single output regression $y = \beta_0 + \beta_1 x$, the parameters (β_0, β_1) refer to

A: $(y - intercept, slope)$	C: $(x - intercept, correlation)$
B: $(x - intercept, MSE)$	D: (pca_1, pca_2)

Q18: Given that the correlation of the input variable and output variable is positive, then the regression equation has

A: a positive intercept.	C: a positive slope.
B: a negative intercept.	D: a negative slope.

Q19: You can discover Overfitting when the model has

A: great results in training and poor results in testing.	C: great results in both training and testing.
B: poor results in both training and testing.	D: poor results in training and great results in testing.

Q20: The accuracy of a model is evaluated using

A:	the training set.	**C:**	the testing set.
B:	the validating set.	**D:**	any of the above.

Q21: Ridge and Lasso regressions are

A:	used for regularization purposes.	**C:**	used to reduce the error.
B:	used for reducing coefficients.	**D:**	All of the above.

Q22: The Grand Total is equal to

a) TP+FN b) TP+TAP c) FP+TN d) TAP+TAN

Q23: TP+FN =

a) TAP b) TAN d) TPP d) TPN

Q24: Which of the measures combines Precision and Sensitivity in a single measure?

a) Accuracy b) Sensitivity c) F_β score d) SSE

Q25: F_1 score considers

a) Precision b) Recall c) Sensitivity d) Precision and Recall

Q26: Given: TAP=70, TAN=130, then the Accuracy of All_Negative_Model is

a) 70% b) 40% c) 35% d) 65%

Q27: Given: TP=10, TN=20, GT=120, then the accuracy =

b) 25% b) 30% c) 20% d) 10%

Q28: Use the data-driven cost matrix to calculate the profit per record.

a) 3$ b) 6$ c) 8$ d) 9$

TN=15 cost=0$	FP=25 cost=4$
FN=10 cost=10$	TP=50 profit=20$

Q29: Given: FP=10, FN=20, TAP=60, and TAN=40, then Specificity×Recall =

a) 75% b) 66.7% c) 50% d) 24.6%

Q30: Given: TP=30, FP=20, TN=40, and FN=10, then Precision of All_Positive_Model is

a) 10% b) 20% c) 30% d) 50