Machine Learning

Q1) A – Least Square error

Q2) – A – Linear regression is sensitive to outliers

Q3) B – Negative

Q4) B- corrélation

Q5) C -High variance and low bias

Q6) B - Predictive

Q7) D – Regularization

Q8) D – SMOTE

Q9) A – TPR and FPR

Q10) B – false

Q11) B – Apply PCA to project high dimensional data

Q12) A, B and C

Q13) regularization is a technique used in ML to prevent the overfitting of data, overfitting is a phenomenon where model is limited to training dataset & performs exceptionally well in training dataset but unable to perform in test dataset

Q14) Popular algorithm that are used for regularization are lasso regression and ridge regression

Q15) Difference between expected value and actual value are called as error in linear regression equation

Statistics

Q1) A – True

Q2) – A – Central Limit theorem

Q3) – B – Modeling bounded count data

Q4) – D - The square of a standard normal random variable follows what is called chi-squared distribution

Q5) – C - Poisson

Q6) – B - False

Q7) b – Hypothèses testing

Q8) A - 0

Q9) C -Outliers cannot conform to the regression relationship

Q10) normal distribution is a probability distribution where the data near the mean have more chance of occurrence as compared to those who are far from mean. When you plot the normal distribution, it resembles a bell-shaped curve

Q11) there are different ways of handling missing data for example if the data is too big and missing values are relatively low which can remove the missing data.

Or we can use various imputing values based on the data we need to fill. For example, if the data we want to fill is a continuous data we can use mean to impute missing values and if the data is ordinal in nature, we can use mode

Q12) mean imputation might not be the best practise, because filling the null values with mean might change the correlation between independent and dependent variable

Q14) linear regression is a technique to predict the future values of a dependent variable based on the independent variable. it can be done by determining the correlation between dependent and independent variable

Q15) the two main branches of statistics are – Descriptive statistics (ex central tendencies like mean median mode) and inferential statistics (ex regression and correlation analysis)