Linear Regression Subjective Questions

Assignment Related

1Q.)From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?

Ans:Not all variables are nessesary as it presents a strong case of multicollinearity.
For ex, holiday can be eliminated as there is already weekday,

2Q.) Why is it important to use drop_first=True during dummy variable creation?

Ans:drop_first=True is important, as it helps in reducing the extra column created during dummy variable creation. Hence it reduces the correlations created among dummy variables.

3Q.)Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?

Ans:atemp

4Q.)How did you validate the assumptions of Linear Regression after building the model on the training set?

Ans:
-Multivariate normality
-No or little multicollinearity
-No auto-correlation.
-Homoscedasticity.

5Q.)Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?

Ans:
-atemp
-year
-season

Linear Regression Subjective Questions

General

1Q.) Explain the linear regression algorithm in detail.

Ans:Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It's used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories (e.g. cat, dog). There are two main types: Simple regression.

2Q.) Explain the Anscombe's quartet in detail.

Ans : Anscombe's quartet comprises four datasets that have nearly identical simple statistical properties, yet appear very different when graphed. Each dataset consists of eleven (x,y) points.

3Q.) What is Pearson's R?

Ans: Pearson's r is a numerical summary of the strength of the linear association between the variables. If the variables tend to go up and down together, the correlation coefficient will be positive. If the variables tend to go up and down in opposition with low values of one variable associated with high values of the other, the correlation coefficient will be negative.

4Q.) What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?

Ans: Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units.
The terms normalization and standardization are sometimes used interchangeably, but they usually refer to different things. Normalization usually means to scale a variable to have a values between 0 and 1, while standardization transforms data to have a mean of zero and a standard deviation of 1.

5Q.) You might have observed that sometimes the value of VIF is infinite. Why does this happen?

Ans: In general one starts with the selection of all variables, and proceeds by repeatedly deselecting variables showing a high VIF. ... An infinite VIF value indicates that the corresponding variable may be expressed exactly by a linear combination of other variables (which show an **i**nfinite VIF as well).

6Q.)What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.

Ans: The purpose of Q Q plots is to find out if two sets of data come from the same distribution. A 45 degree angle is plotted on the Q Q plot; if the two data sets come from a common distribution, the points will fall on that reference line. ... It's being compared to a set of data on the y-axis,

Linear Regression Subjective Questions