DATE

Linear Regression Q&A

2) It reduces the correlation created during the variable as it deletes the column with high Co-relation.

3) Atemp has the highest Co-relation.

4) I validate the assumption by testing the dataset creating a strategy called train test split.

5) Atem, hum, windspeed

1) The dependent variable are mediumly Co-related with other.

General

1) a) It is a form of predective method which tells us the relation blw Independent and dependent variable.

b) It majorly focuses on the relation to build which can be predected using linear equation.

classmate

PAGE

DATE

2) Anscombe quartet comprises four data sets that have nearly identical simple descriptive statistics, yet have very distributions and appear very different when graphed. Each dataset have 11 points.

3) Person's R is the Correlation co-efficient which actually gives the o/p kind of bivariate Correlation as it measures of linear co-relation b/w two sets of data.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

4) When you have lot of independent data in model a lot of them might be on very different scale which might lead a model with very wried Co-efficients that might be difficult to interpret.

Standardizing: The variable are scaled in such a way that there mean is zero & σ is 1

Normalized: This Scaling happens to be a value b/w 0 & 1.

PAGE

DATE

5) The perfect relation b/w the variables give VIF = infinity

6) Q-Q plot is a plot of the quantiles of the first data set against the quantiles of the second data set.

It gives similar distribution interpretation on Y < X values or X > Y values thus this helps linear regression to see a role of feature in the model.

PDF Created Using

Camera Scanner

Easily Scan documents & Generate PDF

https://play.google.com/store/apps/details?id=photo.pdf.maker