linear regression report

1 requirements

Linear regression analysis makes several assumptions. For one, all observations in the data must be independent of each other (e.g., the data should not include more than one observation on any individual/unit). Furthermore, the data should avoid including extreme values since these will skew the results and create a false sense of relationship in the data. In general, linear regression gives more weight to cases that are far from the average. Can you think of any examples or datasets in which this might pose an issue

Answer:
When I think about how I implement the algorithm then I think there will be some issues that linear regression have. Those data sets are assumed to have a linear relationship, which means if our data are depended relate to each other, then it won't be smart to use linear regression. For example, the time a human run in 1000 meters. Typically, a healthy young man can run significantly faster compare to an old man. Other factors need to be take care with, like the condition of track, the mood etc...I am just saying the relationship between those two cannot be represented as linear regression. Data must be independent, one has nothing to deal with the other one. For example, the prediction between study time and grades might not be linear, even though people would agree if you put more time to study, your grade should be up. However, if a person receive good grades at the beginning, he —— she might lack of the motivation to study, which means those data are not independent.

• In this discussion post, the argument is posed that there is no ethically neutral statistical method, with specific reference to linear regression. What is the basis for this argument? Do you agree or disagree? Why?

Answer: The basis for this argument is that since each data point represents a person we are weighting people differently, some people's data or behavior will be ignored which won't be fair to them. Constantly weed out outliers will cause these ethical issues as well. I agree with it, because our model needs to fit the majority's benefit, in another words, if we count too much outliers then our linear regression model might not be accurate and will lose the meaning of doing it. To resolve this issue, one thing I can think about is to make strict policy that make sure we cover as many as outliers possible, even though some of them will be missed. And, certainly, this problem won't go away. Moreover, all statistical measures and techniques are just tools, which means we need to apply real factors to make decision. Those outliers could help us develop better treatments but we need to keep a balance between all of these.