Questions and Answers for Regression and Classification with Linear Models

Bingzhang Zhu

October 20, 2021

1 What is the difference between Linear Regression and Logistic Regression?

Linear and Logistic regression are the most basic form of regression which are commonly used. The essential difference between these two is that Logistic regression is used when the dependent variable is binary in nature. In contrast, Linear regression is used when the dependent variable is continuous and nature of the regression line is linear.

BASIS FOR COMPARISON	LINEAR REGRESSION	LOGISTIC REGRESSION
Basic	The data is modelled using a straight line.	The probability of some obtained event is represented as a linear function of a combination of predictor variables.
Linear relationship between dependent and independent variables	Is required	Not required
The independent variable	Could be correlated with each other. (Specially in multiple linear regression)	Should not be correlated with each other (no multicollinearity exist).

Figure 1: Comparison Chart

Reference: https://techdifferences.com/difference-between-linear-and-logistic-regression.html

2 Will the decision boundary be linear or non-linear in logistic regression models?

The decision boundary is essentially a line or a plane that demarcates the boundary between the classes to which linear regression classifies the dependent variables. The shape of the decision boundary will depend entirely on the logistic regression model.

For logistic regression model given by hypothesis function $h_\theta(x) = g(\theta^T x)$ where g is the sigmoid function, if the hypothesis function is $h(x) = g(1 + 2x + 3x)$ then the decision boundary is linear. Alternatively, if $h(x) = g(1 + 2x^2 + 3x^3)$ then the decision boundary is non-linear.

3 In classification problems like logistic regression, classification accuracy alone is not considered a good measure. Why?

Classification accuracy considers both true positives and false positives with equal significance. If this were just another machine learning problem of not too much consequence, this would be acceptable. However, when the problems involve deciding whether to consider a candidate for life-saving treatment, false positives might not be as bad as false negatives. The opposite can also be true in some cases. Therefore, while there is no single best way to evaluate a classifier, accuracy alone may not serve as a good measure.

Reference: https://www.projectpro.io/article/logistic-regression-interview-questions-/