

## Assignment-8 Logistic Regression

Pradeep Paladugula

6/30/2020

1. When might we use a logistic regression? Why might R tell us it can't converge? What might we do?

Logistic regression is an appropriate regression analysis to conduct when the dependent variable is dichotomous. It is a powerful statistical way of modelling a binomial outcome with one or more independent variables (explanatory variables). The concept of convergence is a well-defined mathematical term. It essentially means that "eventually" a sequence of elements get closer and closer to a single value. We call this single value the "limit". The R tell us it can't converge means that having an output that always varies by some amount. It could even diverge, where its output will undergo larger and larger value swings, never approaching a useful result. More precisely, no matter how long you continue, the function value will never settle down within a range of any "final" value. By using K-means clustering method, repeat until we find the convergence value.

2. When should we use Poisson Regression? When should we switch to Quasi-Poisson or negative binomial?

POISSON REGRESSION should be used when we are having discrete data with non-negative integer values that count something (Y-value). It tells us which explanatory variables have a statistically significant effect on the response variable/independent variables (predictors). In other words, it tells us which X-values work on the Y-value. Quasi-Poisson and negative binomial regression models have equal numbers of parameters, and either could be used for overdispersed count data. They often give us similar results, but there can be striking differences in estimating the "effects of covariates".

"Quasi-Poisson regression": Use Poisson regression to model how changes in the independent variables are associated with changes in the counts. The variance of a quasi-Poisson model is a linear function of the mean.

"Negative binomial regression": Poisson regression assumes that the variance equals the mean. When the variance is greater than the mean, your model has overdispersion. A negative binomial model, also known as NB2, can be more appropriate when overdispersion is present. The variance of a negative binomial model is a quadratic function of the mean.

3. What kind of regression would be best if we have a dependent variable with values: "bad", "ok", "good".

Firstly, Categorical variables require special attention in regression analysis because, unlike dichotomous or continuous variables, they cannot be entered into the regression equation just as they are. Ordinal logistic regression is the best suited regression model. Ordinal logistic regression models the relationship between a set of predictors and an ordinal response variable. An ordinal response has at least three groups which have a natural order, such as "bad", "ok", "good"