

Lachlan Sinclair

2/17/2020

Week 6 Assignment 1 Reflections

1. I am slightly confused about what this question is asking. Linear discriminant analysis categorizes instances by using the predictors values (typically normalized values) in the developed linear equation. Then based off whether this returns a negative or positive value it categorizes the instance as one of the binary targets. So, in a generic sense this can fail if the predictors provided do not allow for any means of splitting the dataset into the binary classes in a statistically significant manner. Essentially if no single predictor or combination of predictors can be split the dataset in a way that results in a significant information gain, a logistic regression model using said predictors would not function well. By this I am implying that some combination of the predictors must be able to split the classes in two groups with a single straight line. Which differs from tree models since they are able to use the predictors to look for multiple groupings of the classes rather than just two. This is a very generalized answer, but I couldn't come up with a more specific meaning to the question.
2. Predictor importance is the information gain achieved by making a decision based off of specifically one attribute, or according to ibm.com "the strength of its relationship to the specified target, independent of other inputs". The weights of the attributes in a LR multilinear function need correctly scale that attributes value to influence the result of entire equation, in conjunction with all of the other attributes and their weights. The key here is that the coefficients in the LR multilinear equation interact with the coefficients (and often normalized) numerical values of the all the other attributes simultaneously. The predictor importance just determines which attribute can provide the most information gain when it is the only decision made on the dataset. This is abundantly clear when looking at the screen shots for this assignment, you can see that the attribute with the highest predictor importance has a very low magnitude coefficient when compared to others (for example the mean area). However, interestingly enough on page 87 of the textbook it mentions that "these weights are often loosely interpreted as importance indicators of the features". So, once the LR function has been developed people do sometimes make inferences in the opposite direction posed in the question, however I think the key work in that quote is loosely.