Logistic Regression

Prof. Eric A. Suess

Feburary 15, 2021

Introduction

Lantz Chapter 6 is about Regression Methods for Predicting Numeric Data. The author presents the basics of Simple Linear Regression and Multiple Linear Regression.

Regression Methods can also be used for Classification.

- Logistic Regression
- CART

Why not linear regression?

The author applies linear regression to the launch data. In this data set the dependent variable is **distress_ct**. This variable has only 3 categories.

And making predictions less than 0 or greater than 3 does not make much sense.

Logistics Regression or **Multinomial Regression** would make more sense.

These are **Generalized Linear Models** (GLMs).

An excellent introduction to Logistic Regression

Here is a link to an R-bloggers post How to perform a Logistic Regression in R.

The author of the post creates training and test data sets. And introduces the use of the **ROC** to evaluate the fitted model.

Logistic Regression

A logistic regression model, models a **binary dependent variable**

$Y = 1$ or Yes

or

$Y = 0$ or No

where $P(Y = 1|X)$ is modeled in terms of the predictors X.

Logistic Regression

Try

$$P(Y = 1|X) = \beta_0 + \beta_1 X$$

but all probabilities need to be between 0 and 1.

What is used is the **logit** function, to keep the values of the probabilities between 0 and 1.

$$P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

Logistic Regression

So it turns out that the **log odds** are linear

$log \left(\frac{P(Y=1|X)}{1-P(Y=1|X)} \right) = \beta_0 + \beta_1 X$

This gives a nonlinear model that is estimated using MLEs by numerical methods.

Multiple Logistic Regression

Multiple Logistic Regression can be used when there is more than one predictor variable.

Categorical or Numeric variables can be used as predictors.

Evaluations

The **AIC** is used to compare models.

The **ROC** curve is used to compare models.

The Area under the ROC is commonly used to evaluate and compare models.

Logistic Regression

Try Logistic Regression with the **launch** data and the **credit** data.

CART

Try CART with the **credit** data.

This will be using the "C" in CART.

An excellent introduction to Generalized Linear Models (GMLs)

Here is a link to a Quick-R post Generalized Linear Models.