

FALL 2018



APPLIED ENGINEERING DATA ANALYSIS, OPTIMIZATION AND VISUALIZATION

Logistic regression

JOSHUA RHODES, PHD

Research Fellow/Adjunct Professor, The University of Texas at Austin

Last time we got a crash course in linear regression

- Sometimes you want to know a binary result, or a yes/no
- Similar question: what is the probability of something happening given other things
 - What is the probability of a person dying from the flu given: age, sex, location, marital status, etc.
 - What is the probability of someone being a Republican (or Democrat) given: age, sex, location, views on issues, etc.

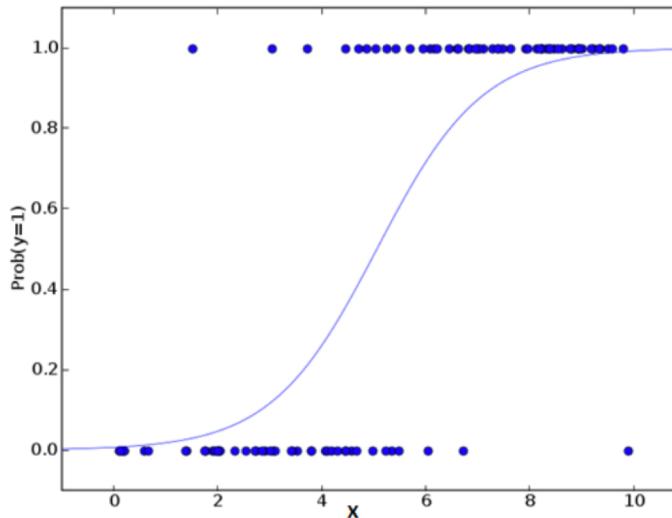
Logistic regression is different than regular regression

- It is not really regression, it is a classification method
- The *log odds* of the outcome is modeled as a linear combination of the predictor variables
 - one of the best mathematical way to replicate a step function (discontinuous function)

We can convert the log odds to a probability

- $\ln(\text{odds}) = \ln(p/(1-p)) = a*x_1 + b*x_2 + \dots + z*x_n$
 - Where a, b, ..., z are our logistic regression coefficients and the x's are our independent variables
- With math...
 - $p = 1/(1 + \exp(-1*(a*x_1 + b*x_2 + \dots + z*x_n)))$

We use logistic regression when we want to estimate the odds of a discrete event



The dots are the actual data and the line is the *predicted probability of $y = 1$, given x*

You can also want the probabilities of more than just a 0 or 1

- Multinomial logistic regression
 - Generalized logit model to more than one value
 - $P(Y_i = 1)$
 - $P(Y_i = 2)$, etc...

Let's look at the probability that a coal plant has retired (or will retire) given age, size, and type...