Logistic regression uses a logistic function to model a binary dependent variable.

To understand logistic regression, we must first understand a few concepts:

**Bernoulli distribution:** probability distribution for experiments with only 2 outcomes (success/failure)

e.g. single coin toss

**success**: P (probability)

**failure**: 1-p

**odds**: p/1-p

**odds ratio:** (p1/1-p1 )/(p2/1-p2)

represents how odds change with 1 unit increase in variable, holding all other variables constant

Logit functions maps the linear combination of variables that result in a domain between 0-1

ln(p/1-p)

The antilog of the logit function helps us graph it as S curve and we get the estimated regression equation

antilog = p/1-p = eB0+B1x1p

A picture containing line chart

Description automatically generated

training data linear model:

Text

Description automatically generated with medium confidence

Multinomial logistic regression

Diagram

Description automatically generated with low confidence

Coefficients become unstable when there is collinearity and when classes are well separated (such as when p> n-1)

Once coefficients estimated, compute

Text

Description automatically generated with medium confidence