Linear Models for Classification Tasks

Ryan Miller

Grinnell College
Statistics

Introduction

Linear regression is a supervised learning approach that models the dependence of a numeric outcome on a set of predictors as linear:

$$Y = w_o + w_1 X_1 + w_2 X_2 + \ldots + w_p X_p + \epsilon$$

- When Y is a binary variable, this model is problematic because predicted values can fall outside of $[0, 1]$

Generalized Linear Models

- **Generalized Linear Models** offer a theoretical framework for adapting the basic structure of linear regression to classification tasks
  - To begin, linear regression can be viewed as the model:

$$y_i \sim N(z_i, \sigma), \text{ where: } z_i = w_o + w_1 x_{i1} + w_2 x_{i2} + \dots$$

- This model has two components:
  - The *linear predictor*, z (called a prediction score by data scientists)
  - A probability model that explains some of the variability in the outcome

Logistic Regression

- The Normal distribution isn't suitable for a binary outcome, but the *Bernoulli distribution* is:

$$Y \sim Ber(g(Z))$$

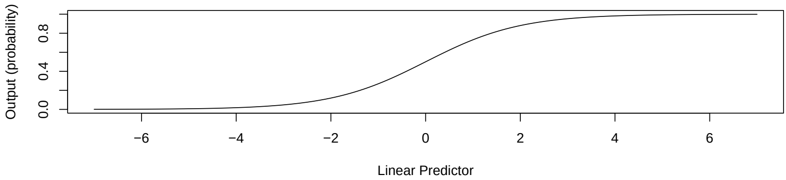
- The mean of a Bernoulli distribution is $Pr(Y = 1)$
  - So, we must transform our linear predictors using a function, $g()$, such that only inputs between 0 and 1 are possible

Logistic Regression

Logistic regression is a generalized linear model that uses the *Bernoulli distribution* and the **sigmoid function**:

$$g(Z) = \frac{1}{1+exp(-Z)}$$

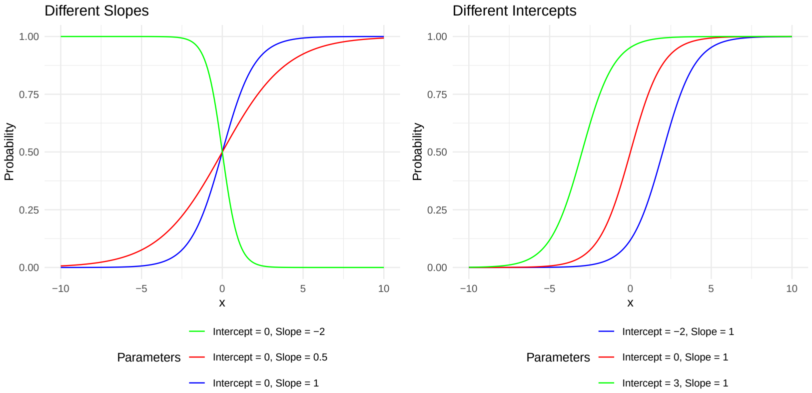
This function maps prediction scores to probabilities, where the observed data (ie: $y_i = 0$ or $y_i = 1$), are
considered samples from a Bernoulli distribution with a mean of $g(Z)$:



Grinnell College
Statistics

Logistic Regression Curves

The shape of the sigmoid curve depends upon the slope ($\hat{w}_1$) and intercept ($\hat{w}_0$):



Grinnell College
Statistics

Logistic Regression (summary)

Putting this all together, logistic regression uses the training data to estimate weights, $\{w_0, w_1, \ldots, w_p\}$, in the model:

$$Pr(Y = 1) = g(Z) = \frac{1}{1+exp(-(w_0+w_1X_1+w_2X_2+...))}$$

We will cover the details of how these weights are estimated in our next unit.

Softmax Regression

- Logistic regression is designed for binary outcomes; however, the method can be generalized to multi-label classification settings
  - **Softmax regression**, also known as multinomial logistic regression, models the probability of class membership for each class via:

$$Pr(y_i = k) = \frac{exp(\mathbf{w}_k^T \mathbf{x}_i)}{\sum_{j=1}^{N_k} exp(\mathbf{w}_j^T \mathbf{x}_i)}$$

- Here N_k is the number of categories
  - Notice the numerator is the exponent of the linear predictor for the category of interest
  - The denominator is the sum of the exponents of the linear predictors for all categories

What to Know for the Next Quiz

- Logistic regression is used to model a binary outcome via the sigmoid function and a linear predictor
  - Softmax (multinomial) regression is used for nomial outcomes
- How the logistic regression (sigmoid) curve looks for various different weights