Logistic Regression

Instructor: Paul M. Magwene

September 30, 2016

Logistic Regression

Logistic regression is used when the dependent variable is discrete (often binary). The explanatory variables may be either continuous or discrete.

Examples:

- whether a gene is turned off (=0) or on (=1) as a function of levels of various proteins
- whether an individual is healthy (=0) or diseased (=1) as a function of various risk factors.
- whether an individual animal died (=0) or survived (=1) some selective event as a function of one or more moprhological traits.

Logistic Regression

Model the binary responses as:

$$P(Y = 1|X_1, \ldots, X_p) = f(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

So we're modeling the probability of the states as a function of a linear combination of the predictor variables.

For logistic regression, f is thus the logistic function:

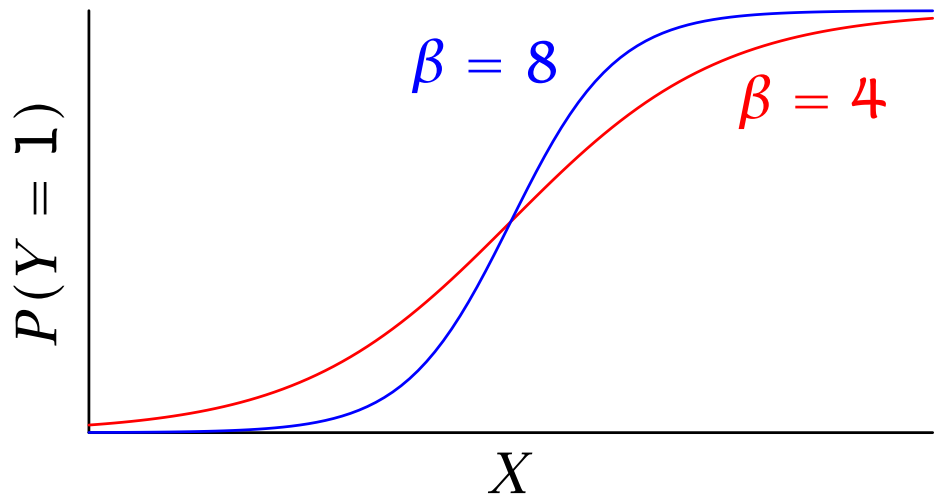
$$f(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

Binary Logistic Regression

Y is a binary response variable (0, 1)

X is a continuous predictor variable

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



Notes on Logistic Regression

- The regression is no longer linear
- Estimating the β in logistic regression is done via maximum likelihood estimation (MLE)

Logistic Regression Example: Sinking of the Titanic