Logistic Regression 101

Bert Gollnick

Logistic Regression

Introduction

- Suitable for classification tasks (don‘t get confused by „regression“)
- Only works for binary classifier
- Independent variables can be continous or discrete
- Related to classical regression

Logistic Regression

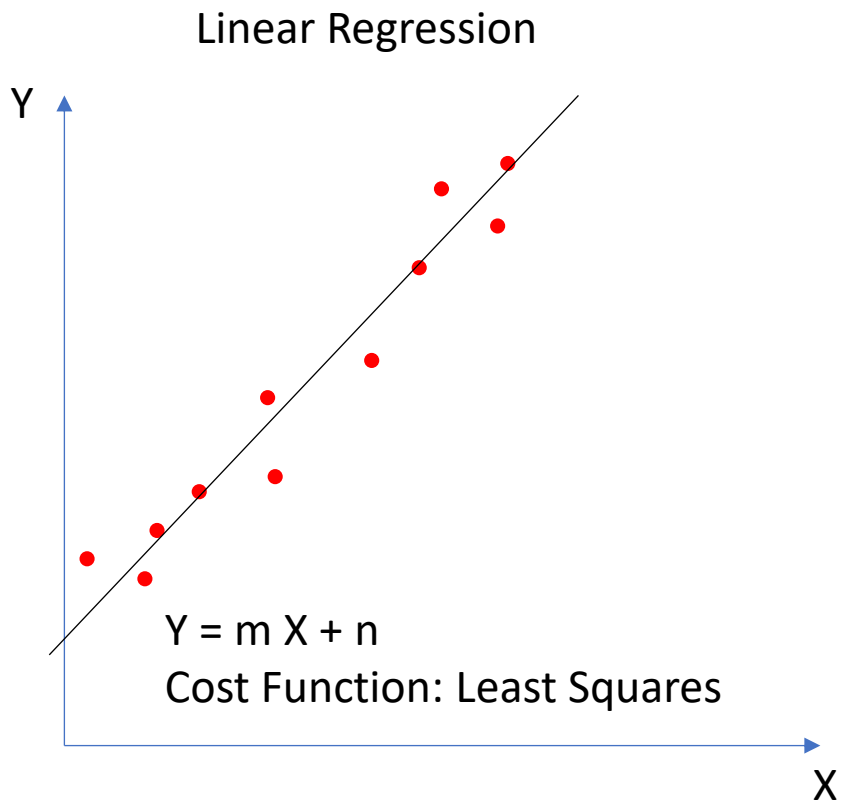
Linear Regression and Logistic Regression

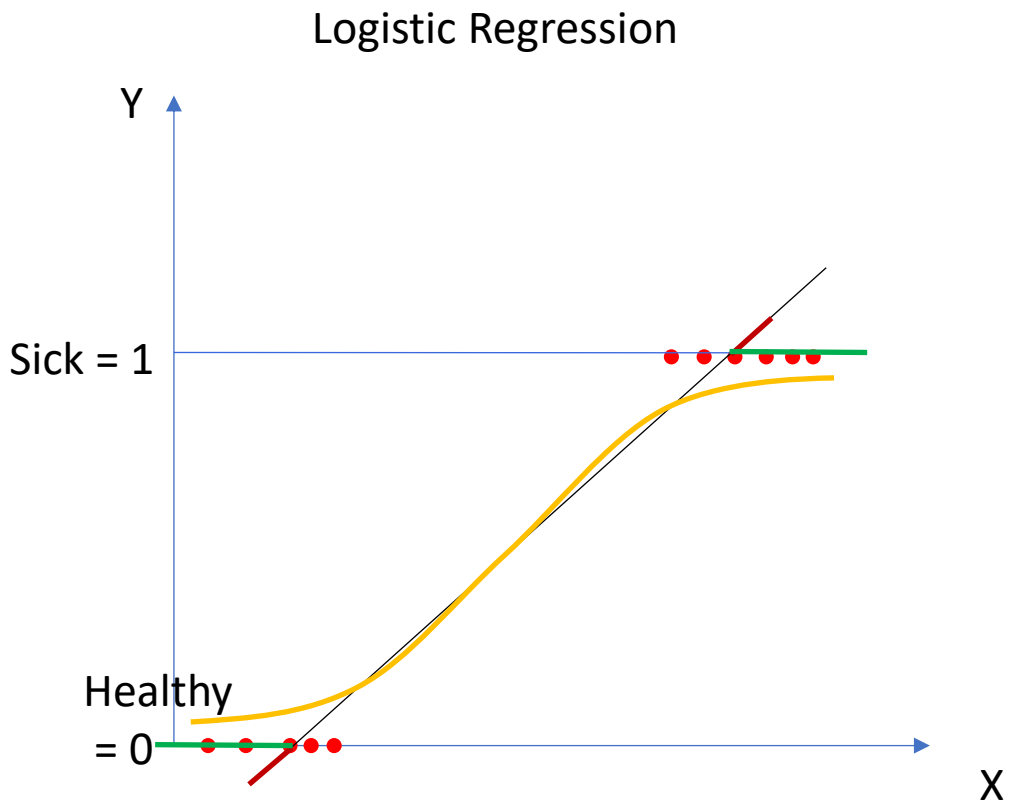
Similarities and Differences

Parameter	Linear Regression	Logistic Regression

Logistic Regression

From Linear Regression to Logistic Regression





Logistic Regression

From Linear Regression to Logistic Regression

Logistic Regression

$$Y = mX + n$$

Transform Target Variable with Sigmoid Function

$$p = \frac{1}{1 + e^{-Y}}$$

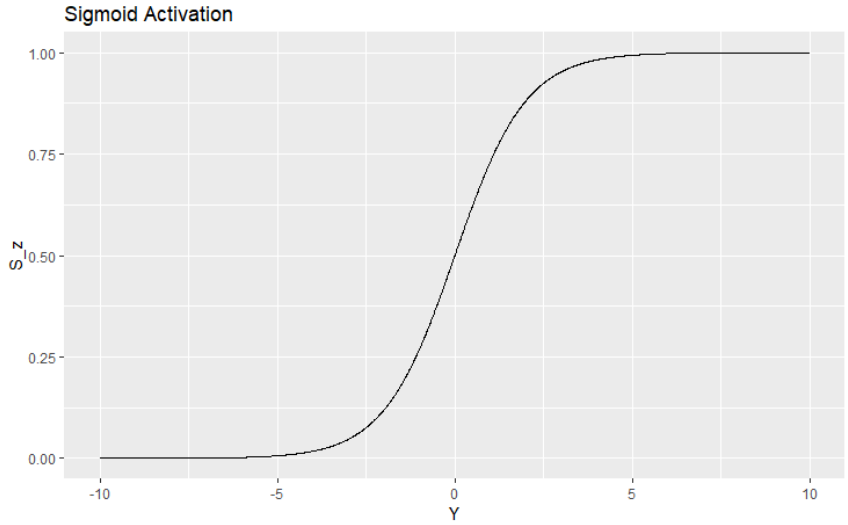
Rewrite Formula:

$$Y = ln\left(\frac{p}{1-p}\right)$$

Logit-Transformation of Target Variable:

$$Y = ln\left(\frac{p}{1-p}\right)$$

$$ln\left(\frac{p}{1-p}\right) = mX + n$$



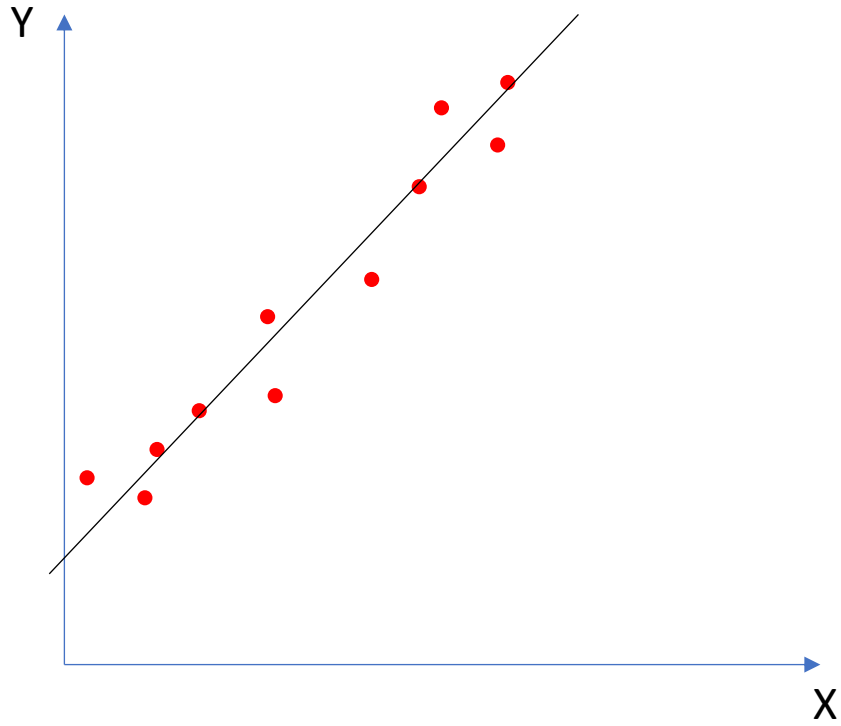
Sigmoid function maps results to 0 to 1 range.

$$S(x) = \frac{1}{1 + e^{-x}}$$

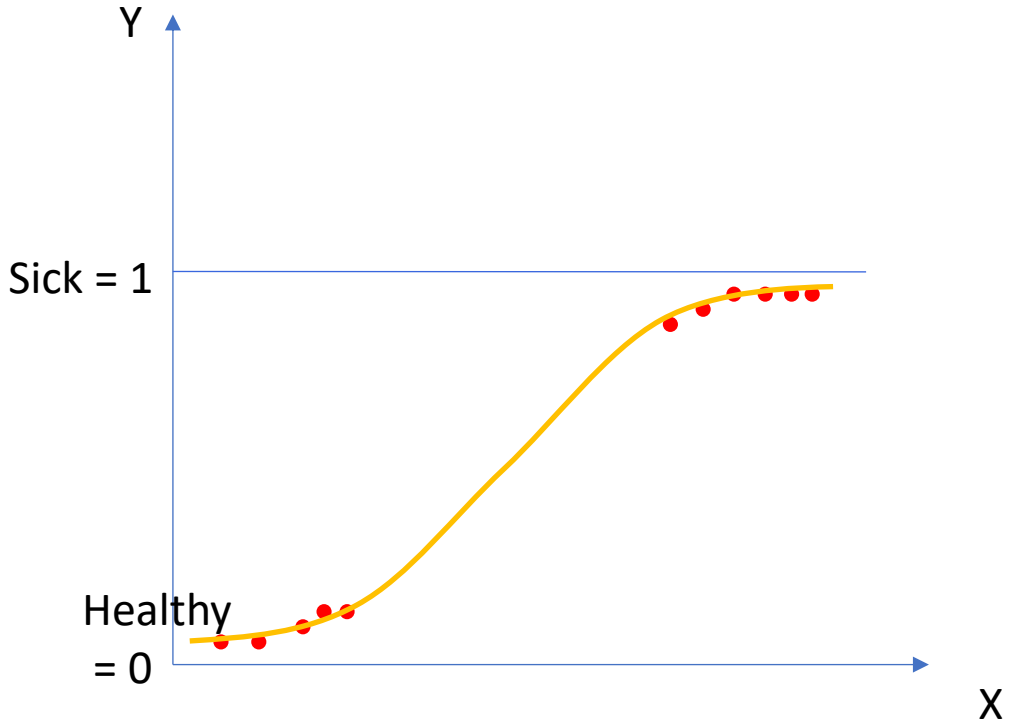
Logistic Regression

From Linear Regression to Logistic Regression

Linear Regression

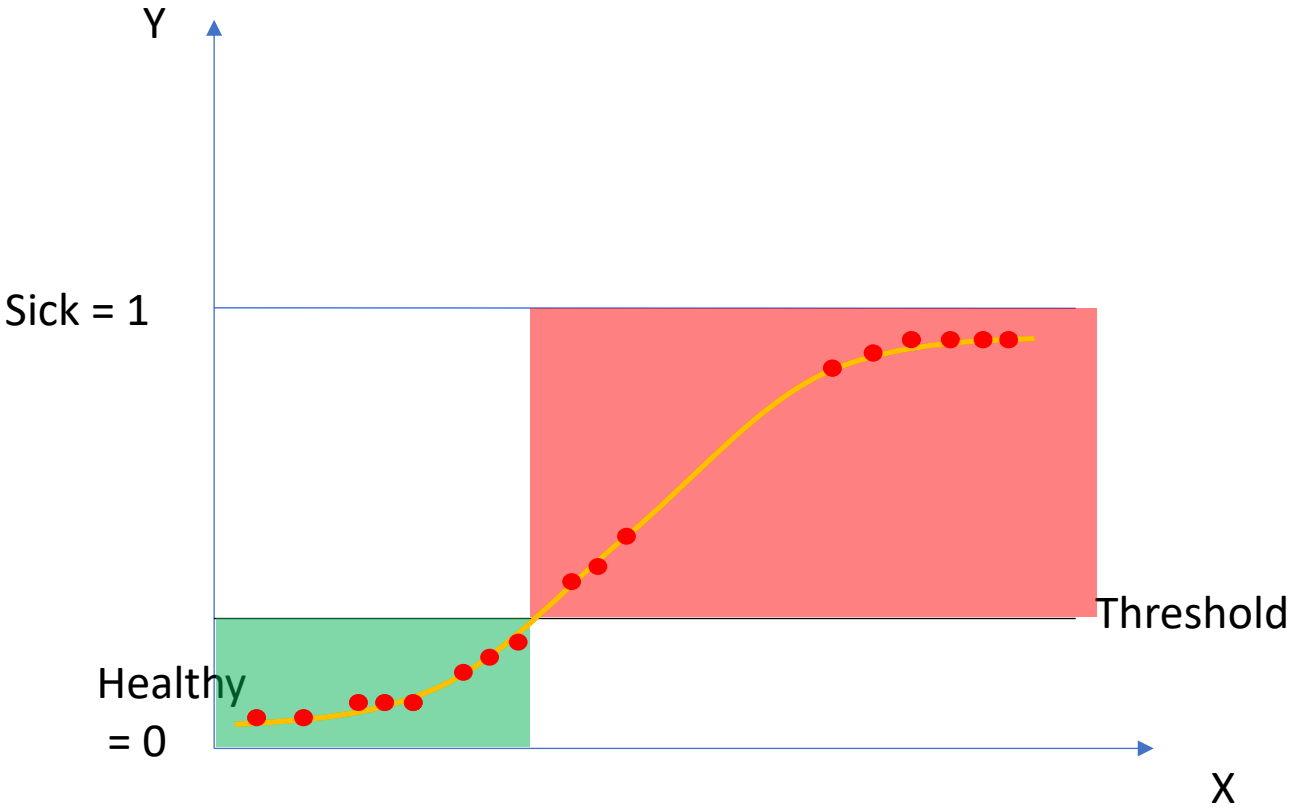


Logistic Regression



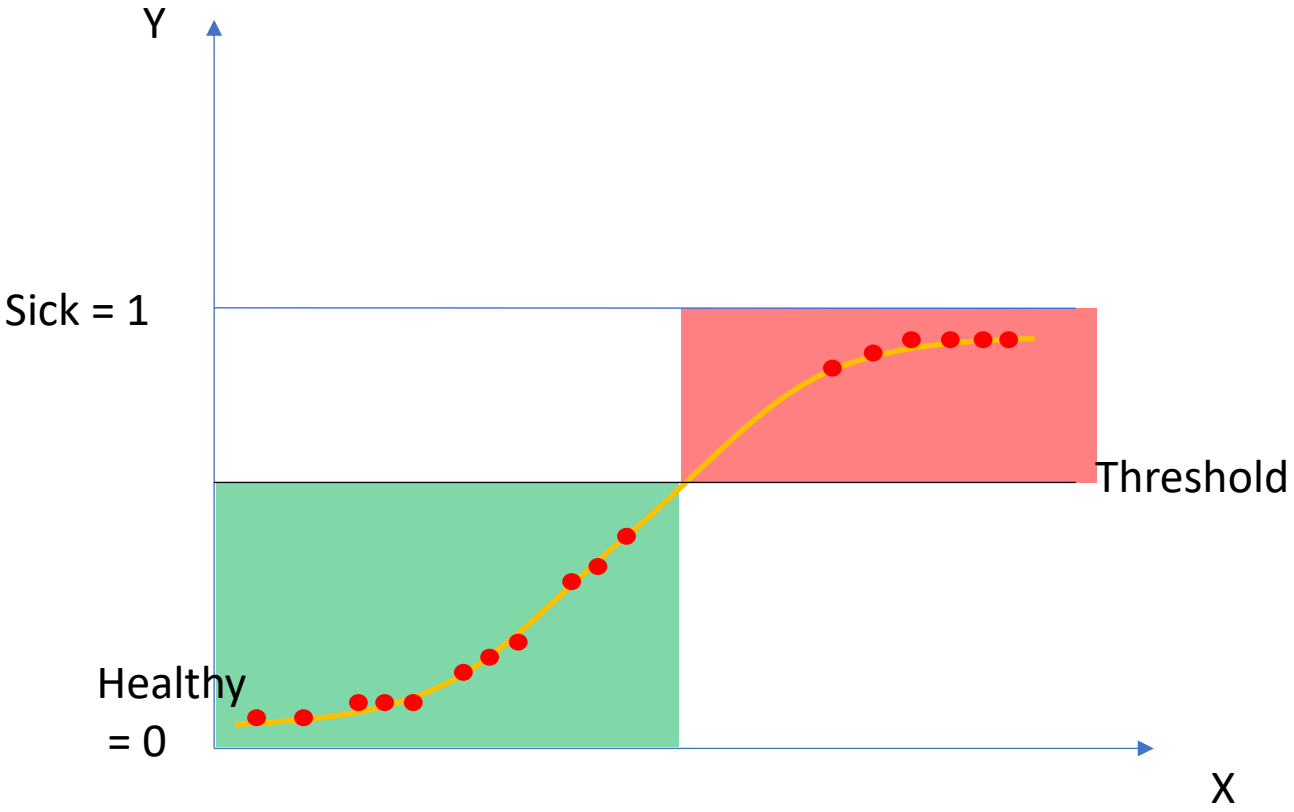
Logistic Regression

From Probabilities to Classes



Logistic Regression

From Probabilities to Classes



Logistic Regression

Advantages / Disadvantages

\+

- Simple to understand
- Low variance
- Can assess variable importance

\-

- Works poor for large number of categorical variables
- Cannot detect complex relationships

Logistic Regression

Classical vs. Bayesian Logistic Regression

- Model Parameters are distributions.
- 95 % Credible Interval defines interval that contains 95 % of the distribution.