Estimating the Logistic Model

Pavlos Protopapas

Lecture Outline

- What is classification
- Classification: Why not Linear Regression?
- Binary Response & Logistic Regression
- Estimating the Simple Logistic Model
- Classification using the Logistic Model
- Multiple Logistic Regression
- Extending the Logistic Model
- Classification Boundaries

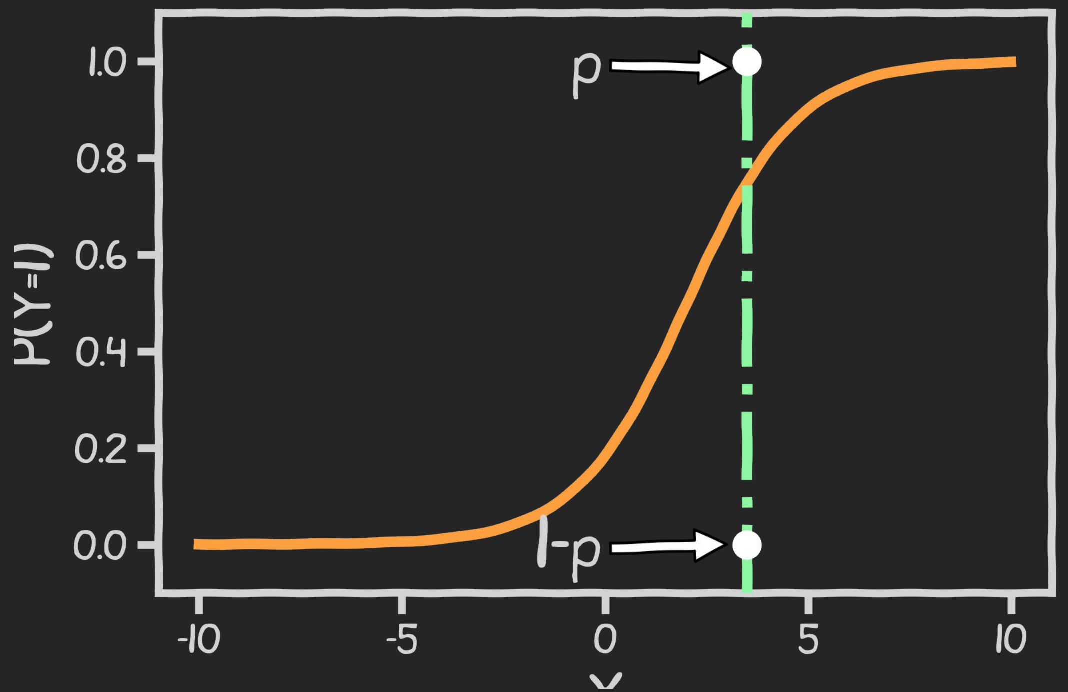
Estimating the Simple Logistic Model

Estimation in Logistic Regression

Unlike in linear regression where there exists a closed-form solution to finding the estimates, $\hat{\beta}_j$'s, that minimize the loss function, logistic regression estimates cannot be calculated through simple matrix multiplication.

Questions:

Estimation in Logistic Regression



Prob. $Y = 1$: $P(Y = 1) = p$
Prob. $Y = 0$: $P(Y = 0) = 1 - p$

$$P(Y = y) = p^y(1 - p)^{(1-y)}$$

where: $p = P(Y = 1|X = x)$ and therefore p depends on X.
Thus not every p is the same for each individual measurement.

Log Regression, tells us the prob of getting 1 or 0 given an x.
For any X the probability of getting heads or tails (1 or 0) is given by with prob given by the logFunc shown in the plot as an orange line./

Likelihood

The likelihood of a single observation for *p* given *x and true label y* is:

$$L(p_i|Y_i) = P(Y_i = y_i) = p_i^{y_i}(1-p_i)^{1-y_i}$$

Given (assuming) the observations are independent, the total prob. (or likelihood) of getting a certain outcome *will be the product of all indiv. prob*

$$L(p|Y) = \prod_i P(Y_i = y_i) = \prod_i p_i^{y_i}(1-p_i)^{1-y_i}$$

The best model is the model that gives us the highest prob. We can adjust the betas such as we get the max prob given our data . We usually do not like to deal with products, (the capital PI) so with a little algebra we can show that finding the max prob is IDENTICAL to finding the min of the –ve log of the prob (or likelihood)
Or minimizing the –ve log likelihood

$$l(p|Y) = -\log L(p|Y) = -\sum_i y_i \log p_i + (1-y_i)\log(1-p_i)$$

This loss fumction is ala as binary cross entropy

Loss Function

$$l(p|Y) = -\sum_i \left[y_i \log \frac{1}{1+e^{-\beta X_i}} + (1-y_i) \log\left(1 - \frac{1}{1+e^{-\beta X_i}}\right)\right]$$

How do we minimize this?

Differentiate wrt to betas, equate to zero and solve for it!

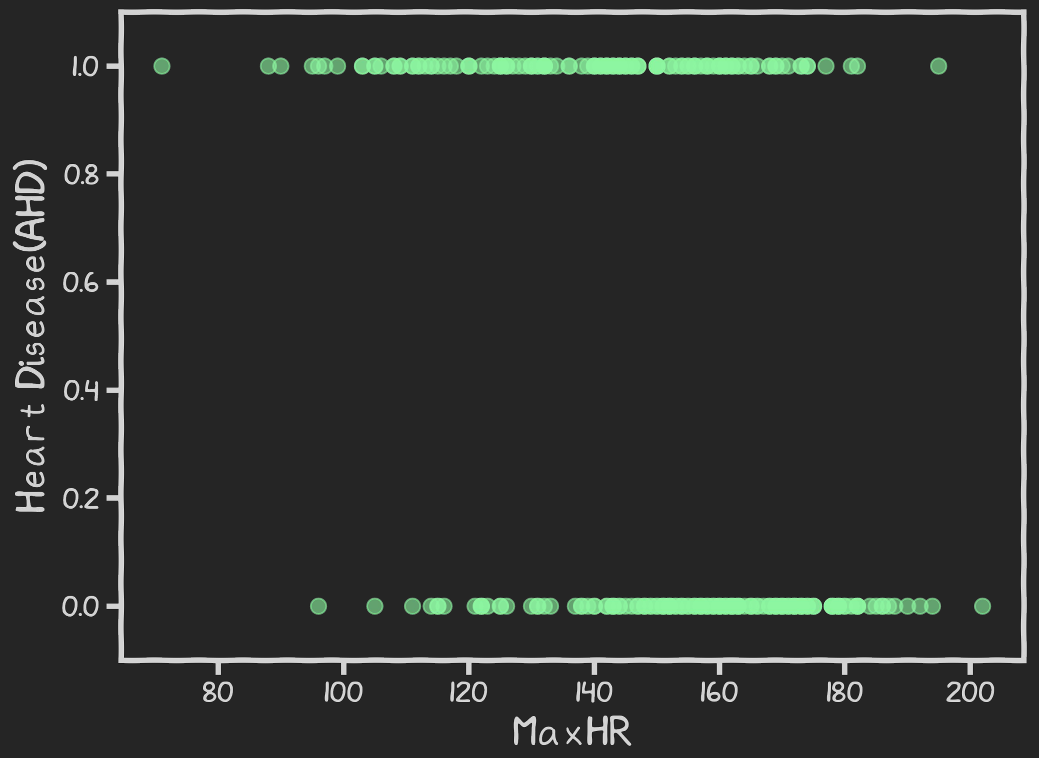
But jeez does this look messy?! And we know it doen not have a closed form solution.

So how do we determine the parameter estimates? Through an iterative approach such as Gradient Descent.

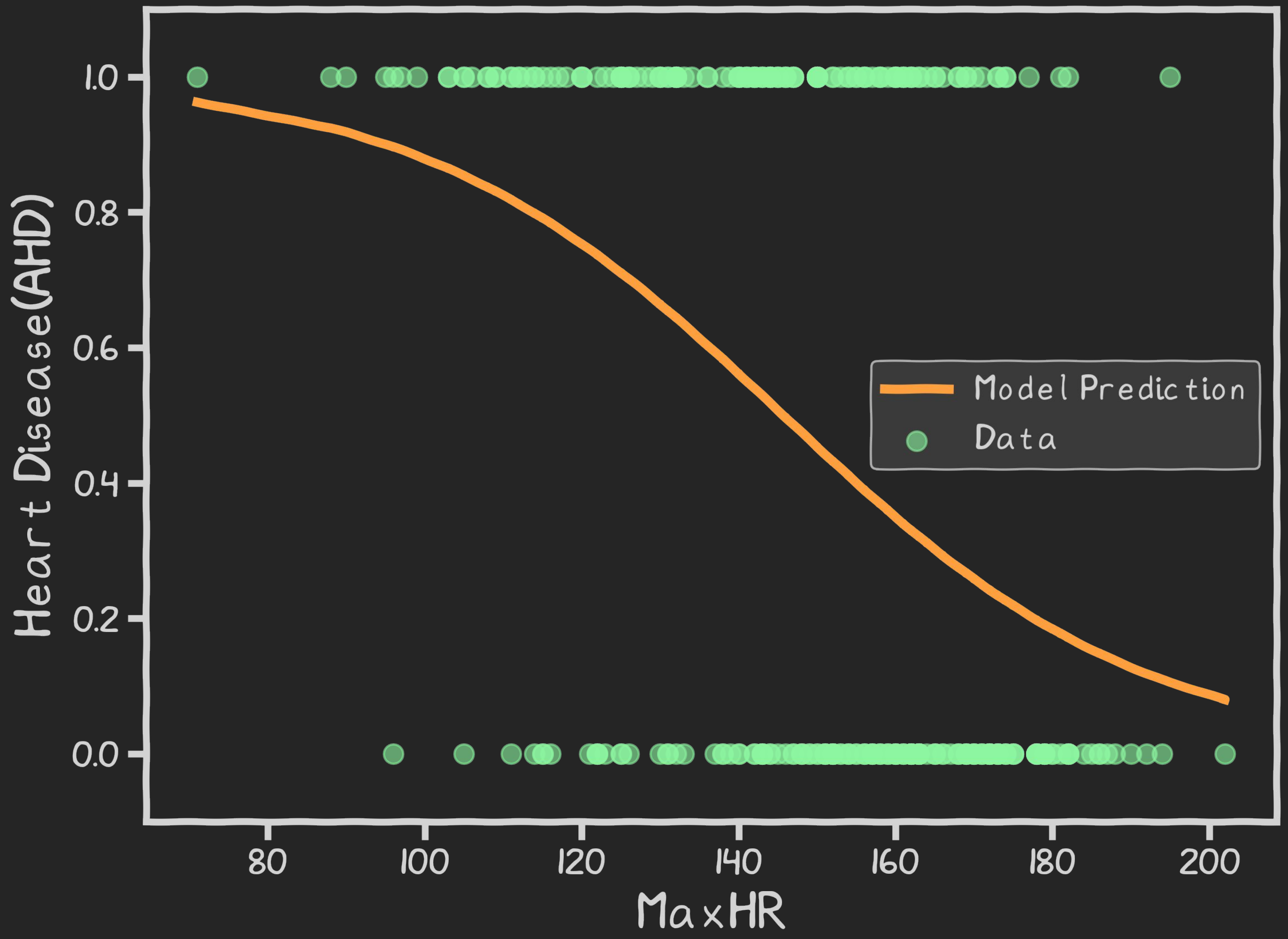
Heart Data: logistic estimation

We'd like to predict whether or not a person has a heart disease. And we'd like to make this prediction, for now, just based on the MaxHR.

~~How should we visualize these data?~~



Heart Data: logistic estimation



Heart Data: logistic estimation