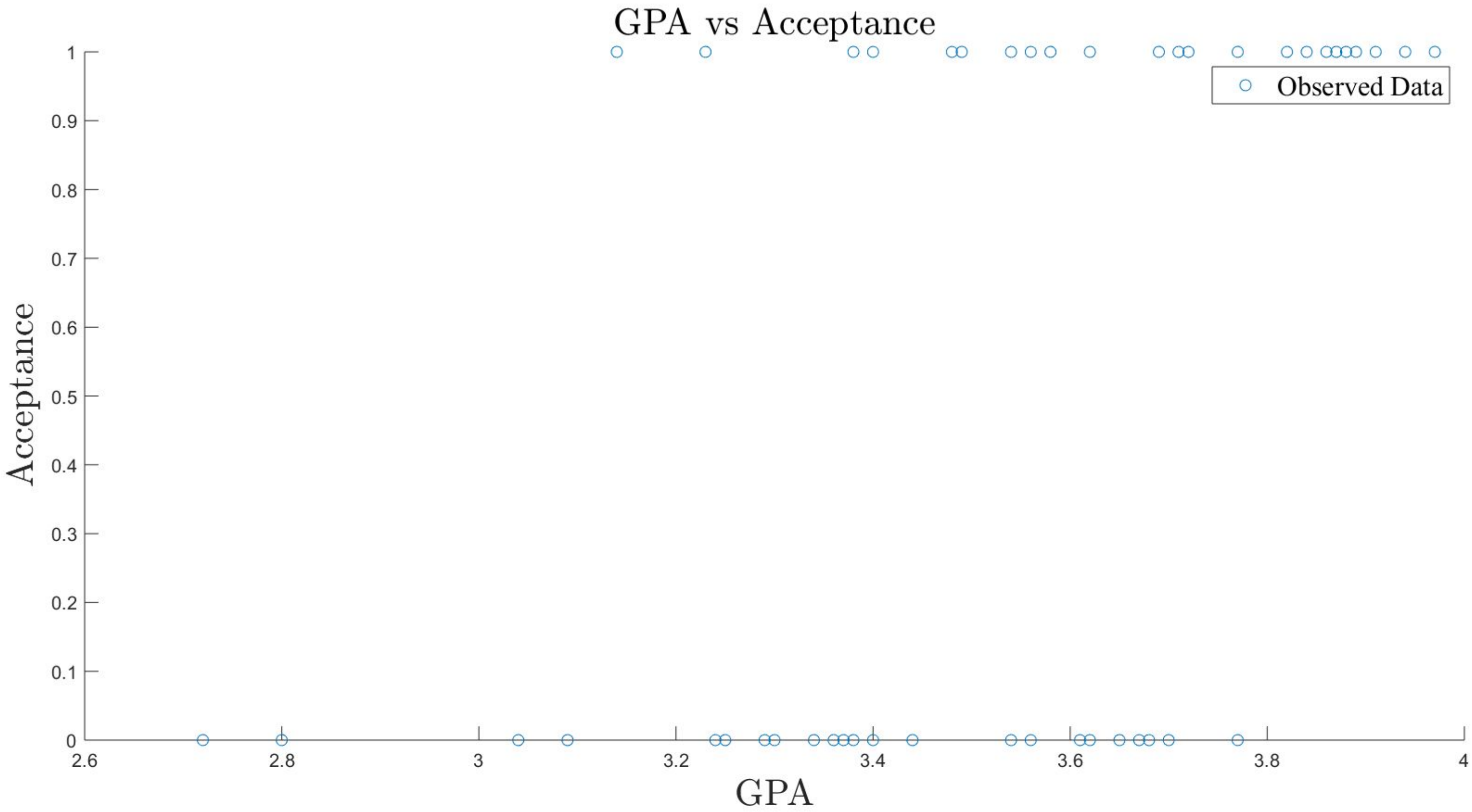
Logistic Regression

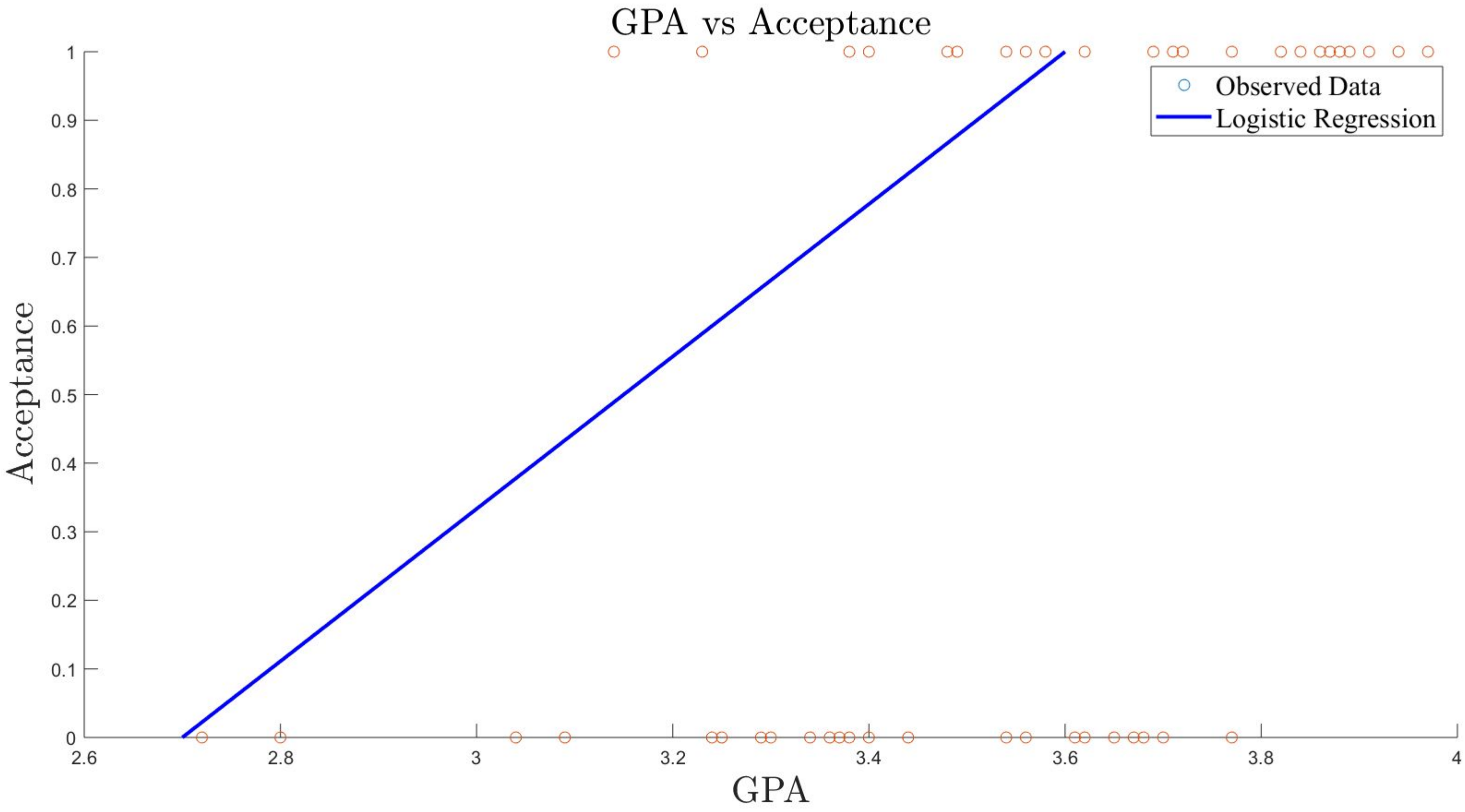
INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

The Problem



GPA vs Acceptance
Observed Data
Acceptance
GPA
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2.6
2.8
3
3.2
3.4
3.6
3.8
4

The Problem


GPA vs Acceptance
Observed Data
Logistic Regression
Acceptance
GPA
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2.6
2.8
3
3.2
3.4
3.6
3.8
4

Motivation

- Logistic regression is the type of regression we use for a binary (or discrete) response variable ($Y \in \{0,1\}$)

- Linear regression is the type of regression we use for a continuous, normally distributed response ($Y \in$ □^{m}) variable

- Use a function to map real numbers to $\{0,1\}$

Dependent Variable Characteristics

- Each trial has two possible outcomes: success or failure.

- The probability of success (call it p) is the same for each trial.

- The trials are independent, meaning the outcome of one trial doesn't influence the outcome of any other trial.

Bernoulli Distribution

$$Pr(y|x;p) = \begin{cases} p, & y = 1 \\ 1 - p, & y = 0 \end{cases}$$

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

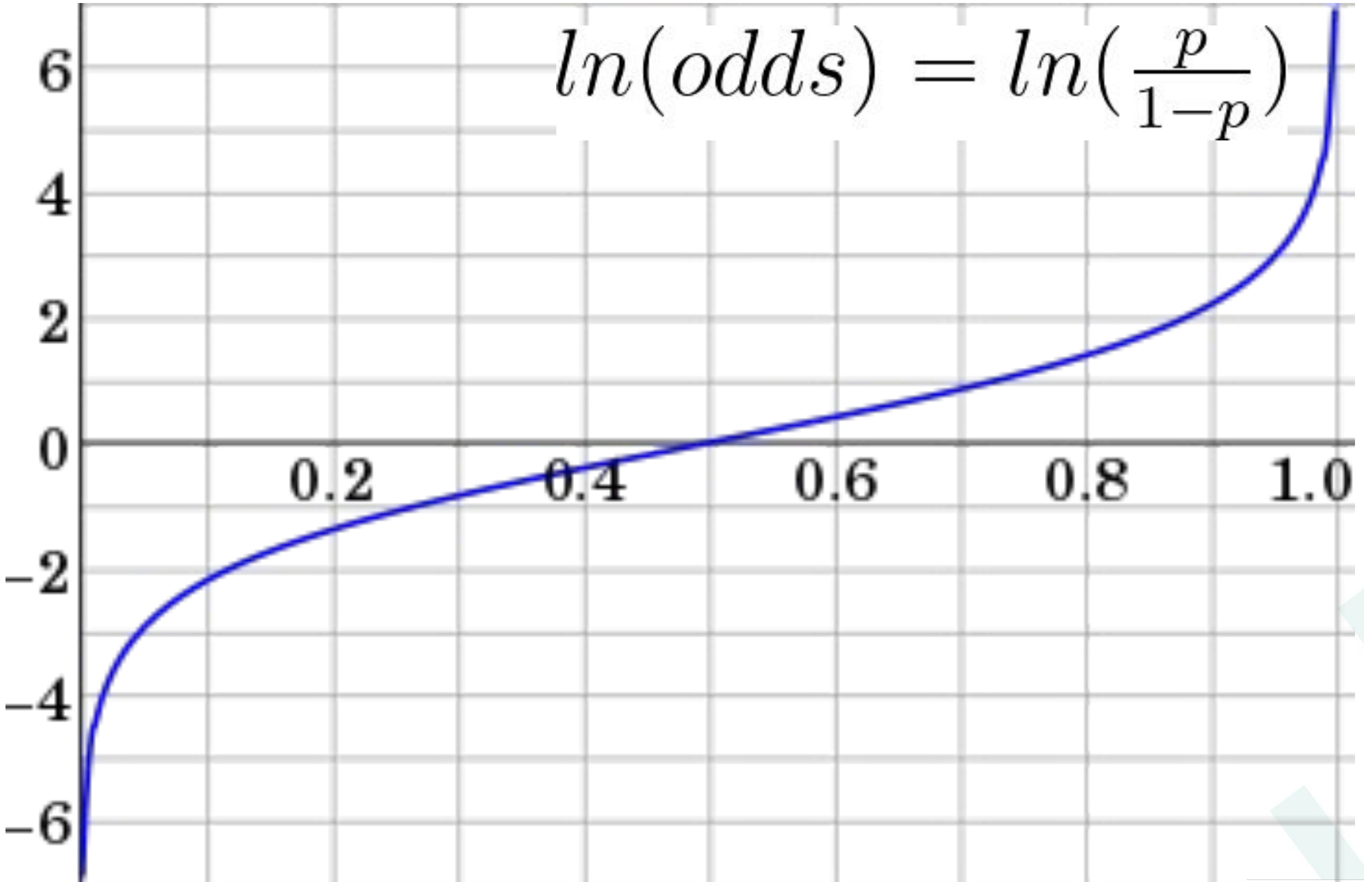
- Input: Linear combination of variables
- Output: Bernoulli distribution *p*

Logit

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

- Range = -inf to +inf
  - Solves the problem we encountered in fitting a linear model to probabilities
  - *P* only range from 0 to 1, we can get linear predictions that are outside of this range

Logit



Inverse Logit

- We want to predict *p*, hence p has to be our Y axis rather X axis.

- The inverse of the logit function is the sigmoid function

- $logit^{-1}(z) = \sigma(z) = 1/(1 + exp(-z))$

Inverse Logit: Derivation

$$logit(p) = log\frac{p}{1-p}$$

Let logit(p) = ŷ:

$$\hat{y} = log\frac{p}{1-p}$$

Taking exponential both the sides:

$$e^{\hat{y}} = \frac{p}{1-p}$$

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

Adding 1 both the sides:

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

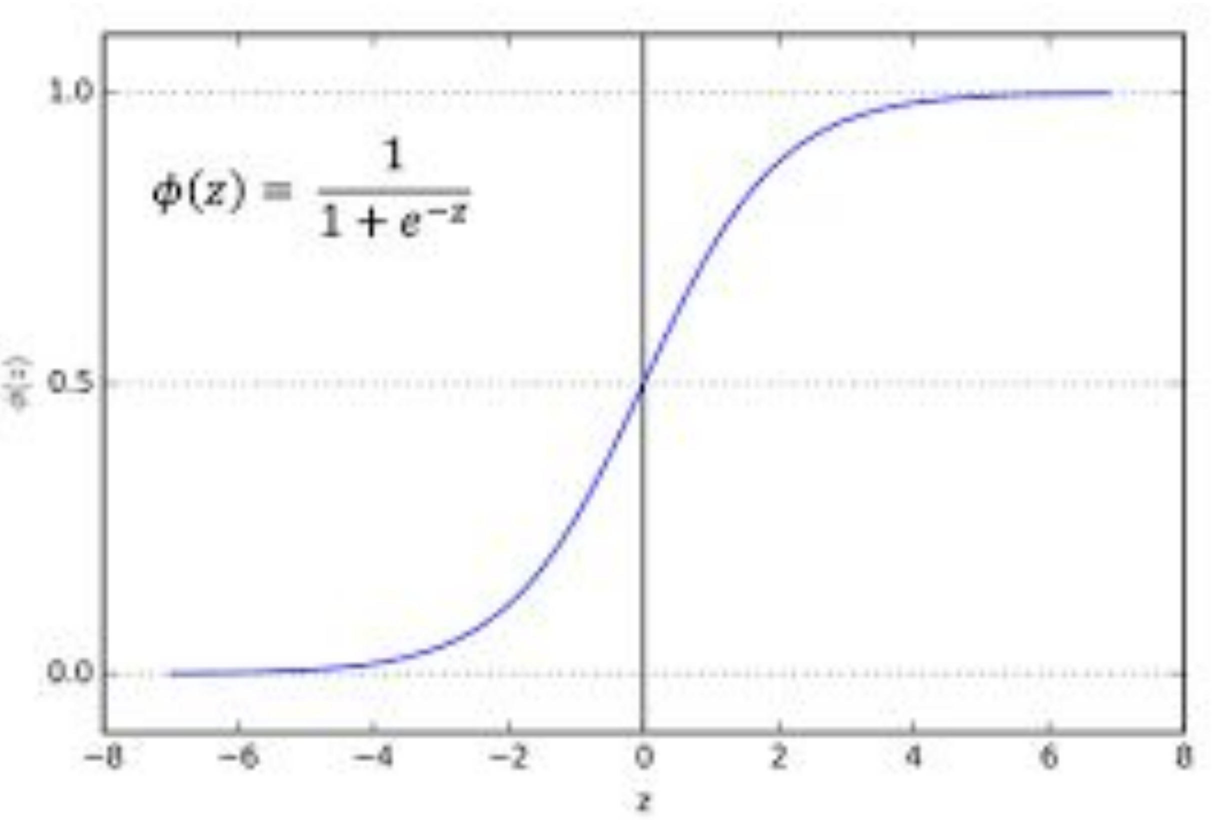
Cross-Multiplication:

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

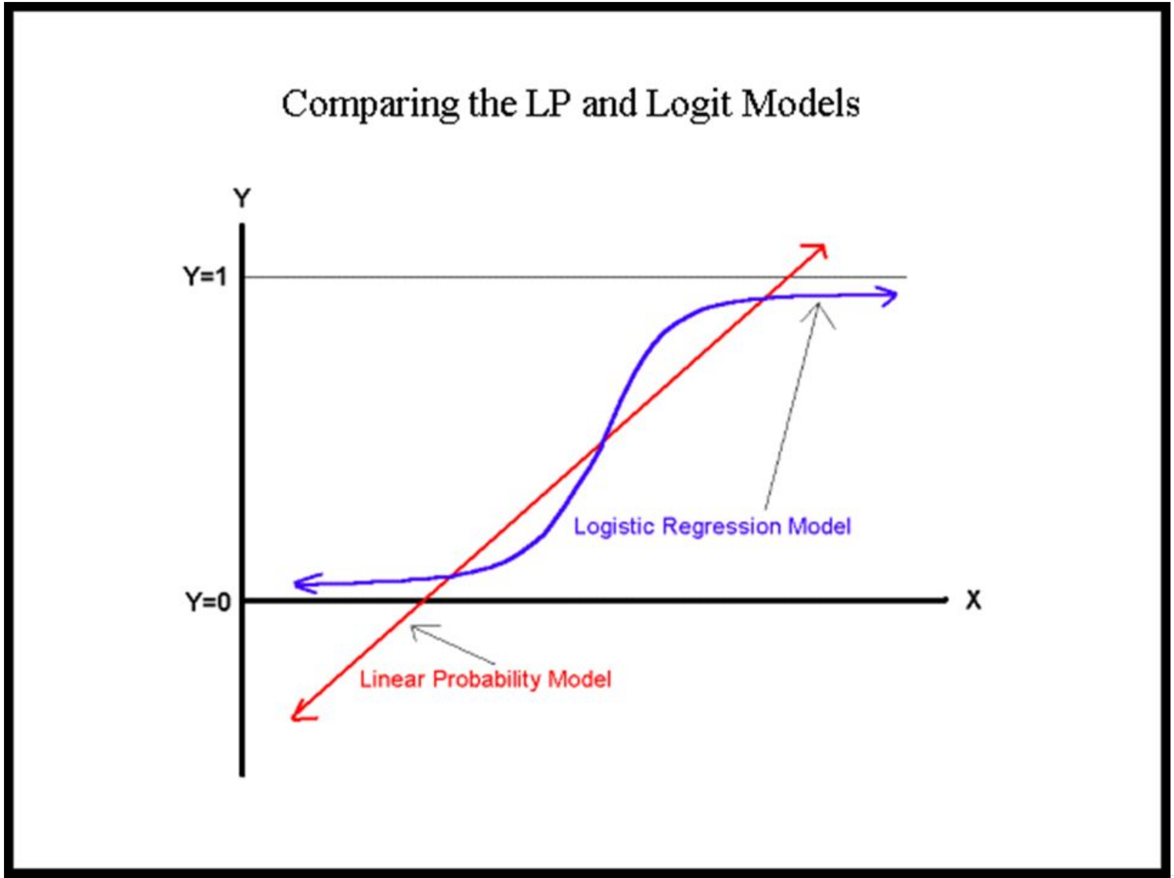
Simplifying it further:

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

Inverse Logit



Linear Regression vs Logistic



Linear Regression vs Logistics Regression

- Linear regression we had
  - $h_\theta(x) = \theta^T x$
- logistic regression we have
  - $h_\theta(x) = 1/(1+e^{-\theta^T x})$

Parameter Estimation: Linear Regression on 'log(odds)'

- Hypothesis (Predicted)

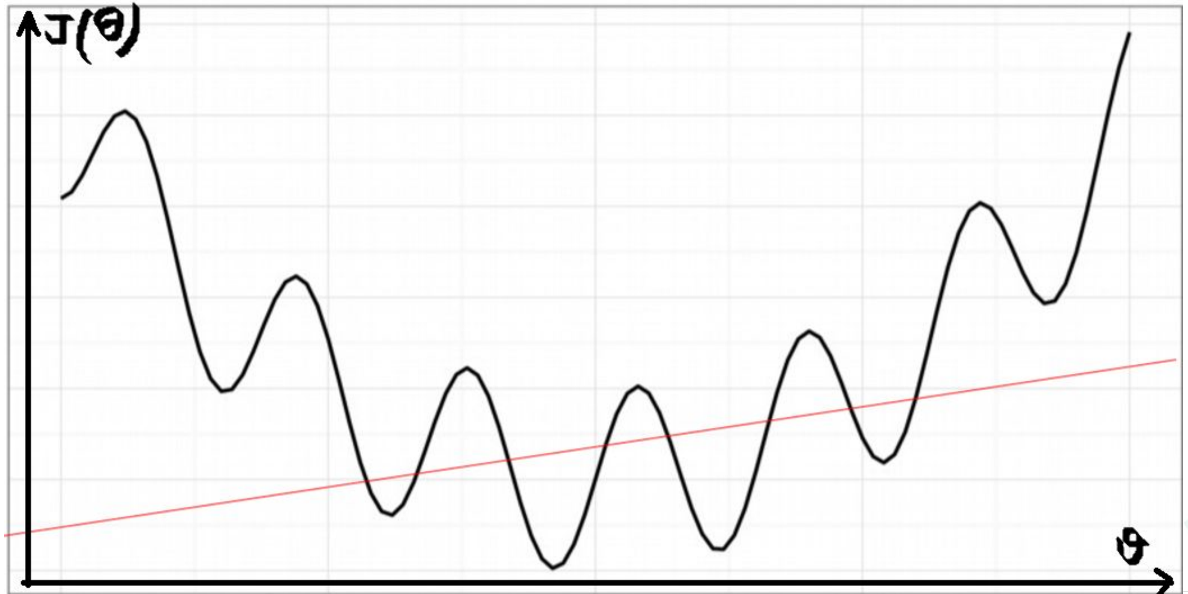
$$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2.$$

- Cost Function

Parameter Estimation: Linear Regression on 'log(odds)'

- Hypothesis $$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

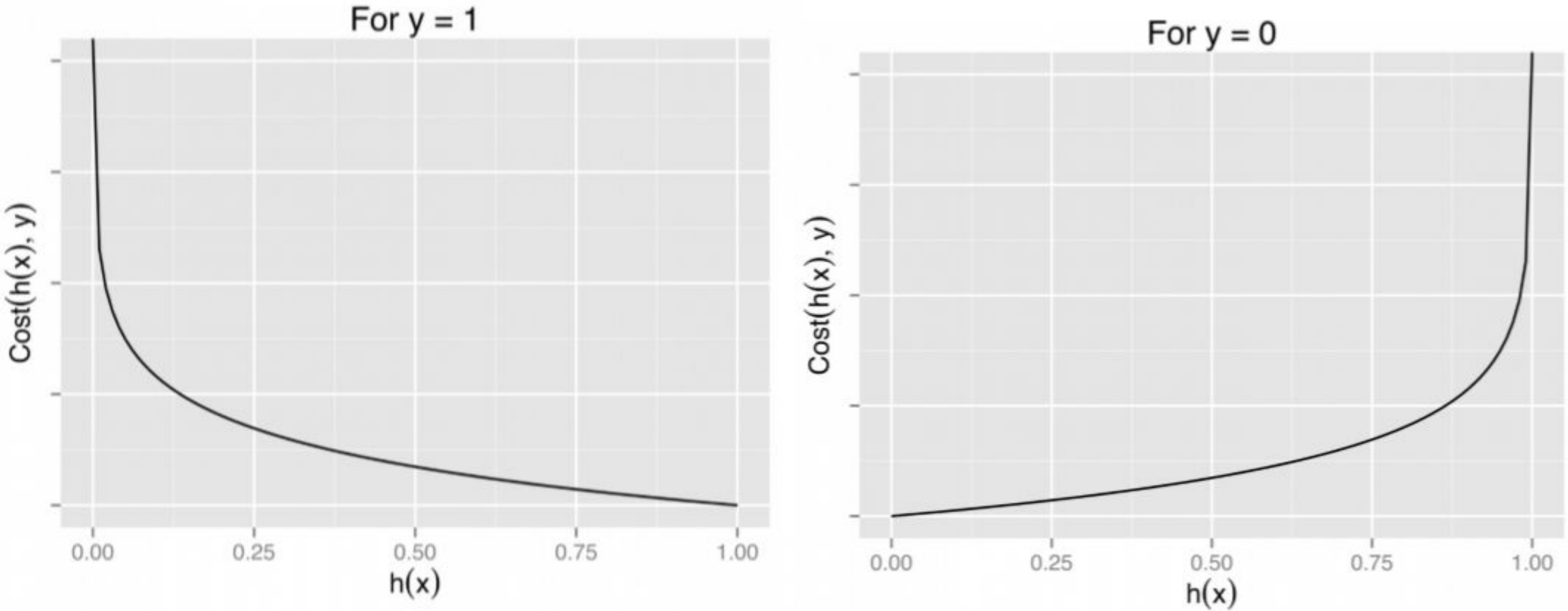


Parameter Estimation: Linear Regression on 'log(odds)'

- Hypothesis $$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

- Cost Function $$\begin{cases} -log(h_\theta(x)) & \text{if } y = 1 \\ -log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Parameter Estimation: Linear Regression on 'log(odds)'



General Cost Function

$$J(\theta) = -\frac{1}{m}\sum\left[y^{(i)}\log(h\theta(x(i))) + \left(1 - y^{(i)}\right)\log(1 - h\theta(x(i)))\right]$$

$$\theta j := \theta j - \alpha\frac{\partial}{\partial\theta j}J(\theta)$$

Repeat $\{$

$$\theta_j := \theta_j - \alpha\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

(simultaneously update all θ_j)

$\}$