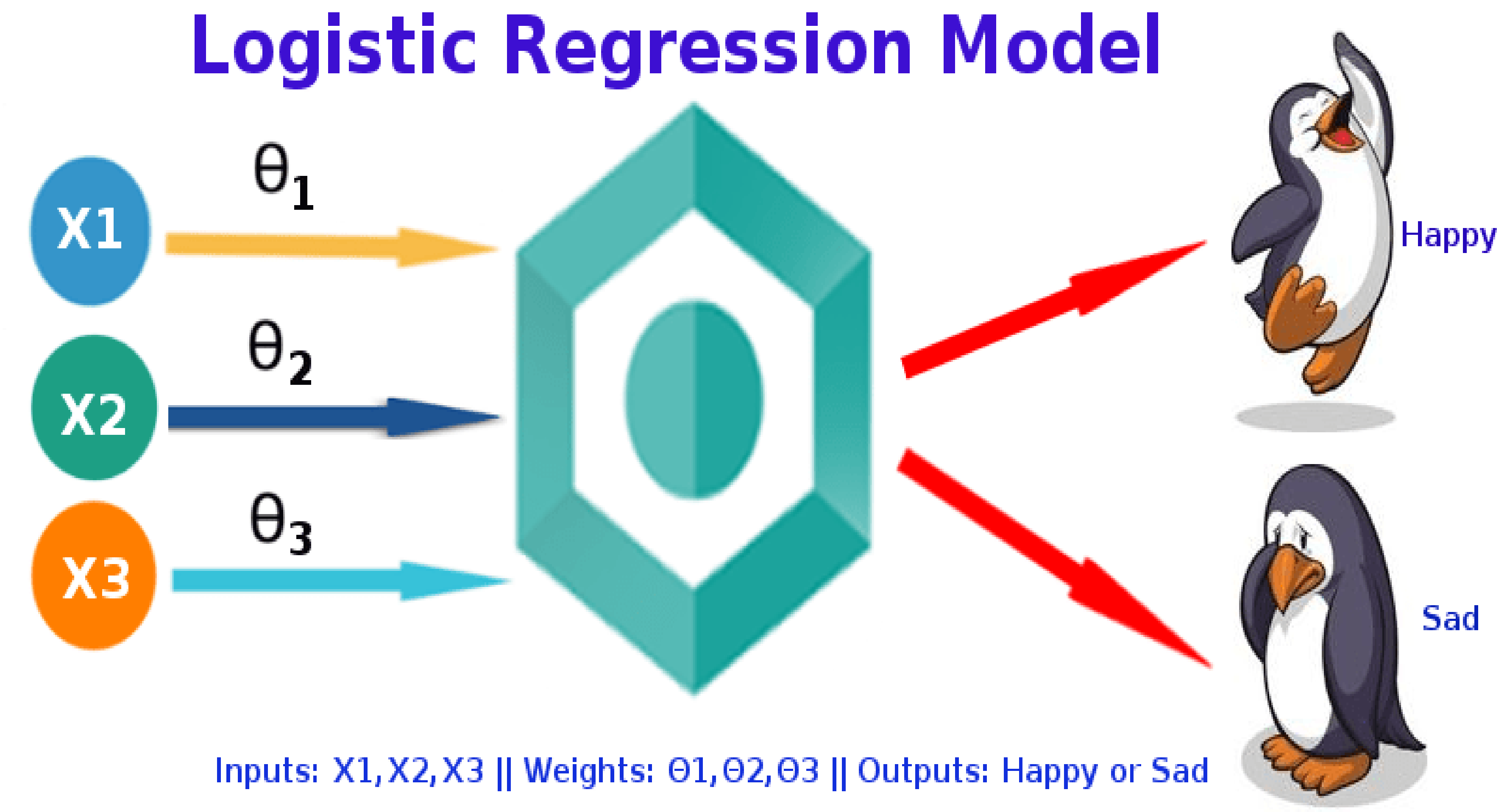
Logistic Regression

Dr. Dinesh Kumar Vishwakarma
Professor,
Department of Information Technology,
Delhi Technological University, Delhi

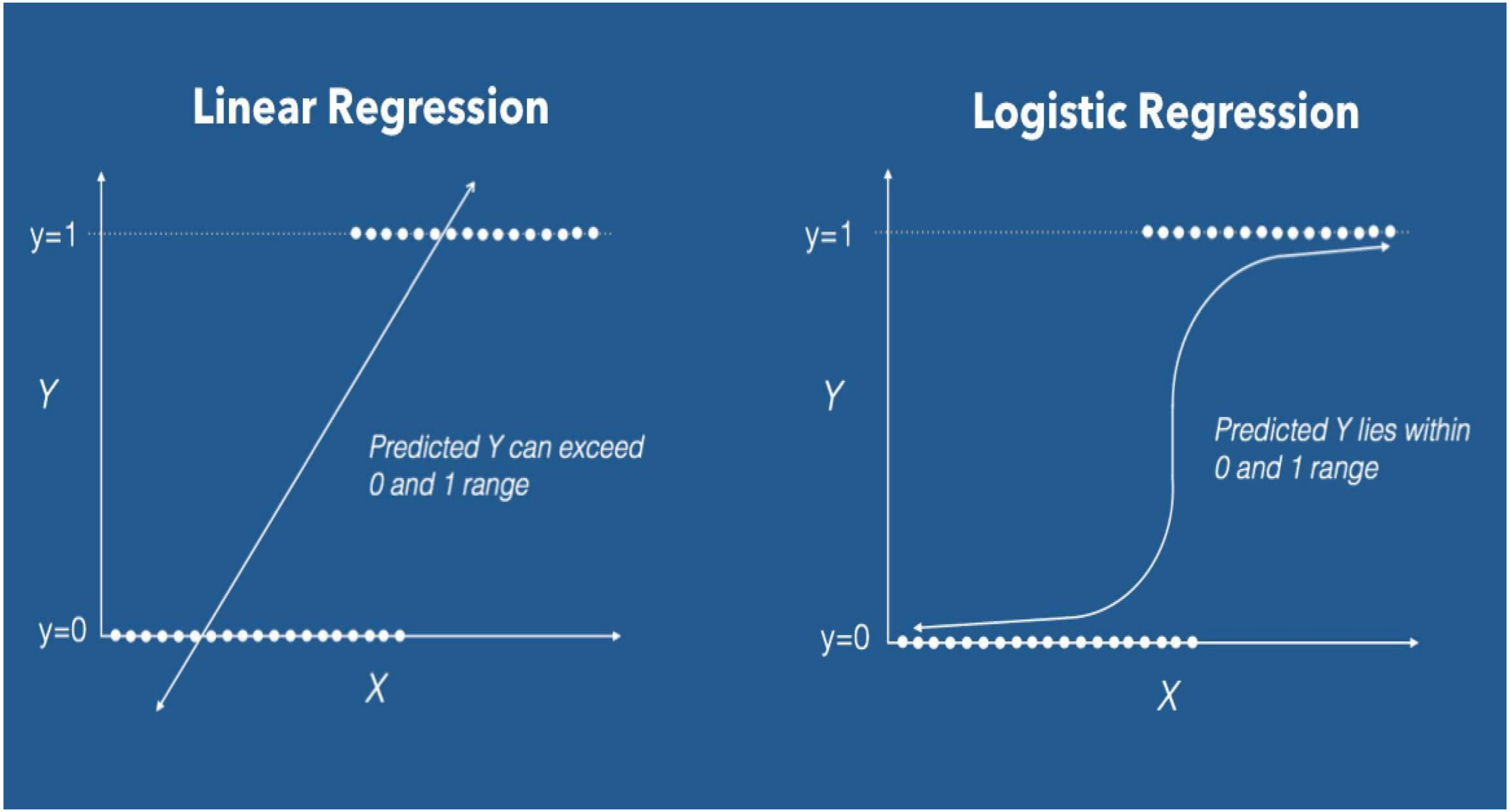
Logistic Regression: Intro

- Logistic regression extends the ideas of linear regression to the situation where the dependent variable, Y , is categorical.
- Now suppose the dependent variable *y* is **binary.**
- It takes on two values "Success" (1) or "Failure" (0)
- We are interested in predicting a *y* from a continuous independent variable *x*.
- This is the situation in which **Logistic Regression** is used.

Logistic Regression



Linear vs Logistic



Example

- Based CGPA of UG, a student will get the admission in PG? Yes/No
- The values of *y* are 1 (Success) or 0 (Failure). The values of *x* range over a continuum. Raining or Not.
- A categorical variable as divides the observations into classes of a stock such as holding /selling / buying, then categorical variable with 3 categories. “hold" class, the “sell" class, and the “buy” class.
- It can be used for classifying a new observation into one of the classes, based on the values of its predictor variables (called “classification").

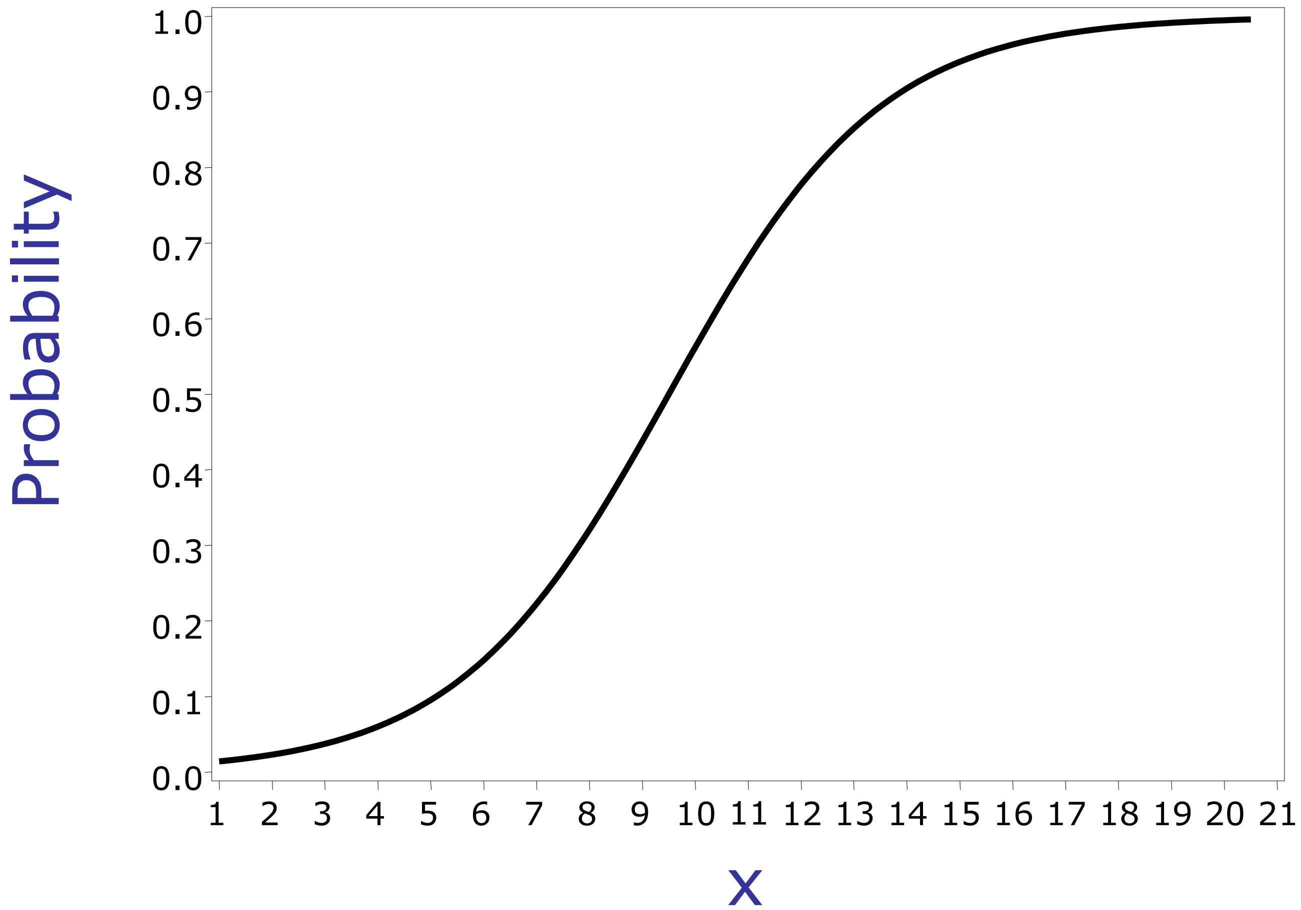
Applications

- Logistic regression is used in applications such as:
  - Classifying customers as returning or non-returning (classification)
  - Finding factors that differentiate between male and female top executives (profiling)
  - Predicting the approval or disapproval of a loan based on information such as credit scores (classification).
- Popular examples of binary response outcomes are
  - success/failure, yes/no, buy/don't buy, default/don't default, and survive/die.
- We code the values of a binary response Y as 0 and 1.

Introduction Logistic Regression

- Most important model for ***categorical response*** (y_i) data
- Categorical response with 2 levels (*binary*: 0 and 1)
- Categorical response with ≥ 3 levels (nominal or ordinal)
- Predictor variables (x_i) can take on *any* form: binary, categorical, and/or continuous.

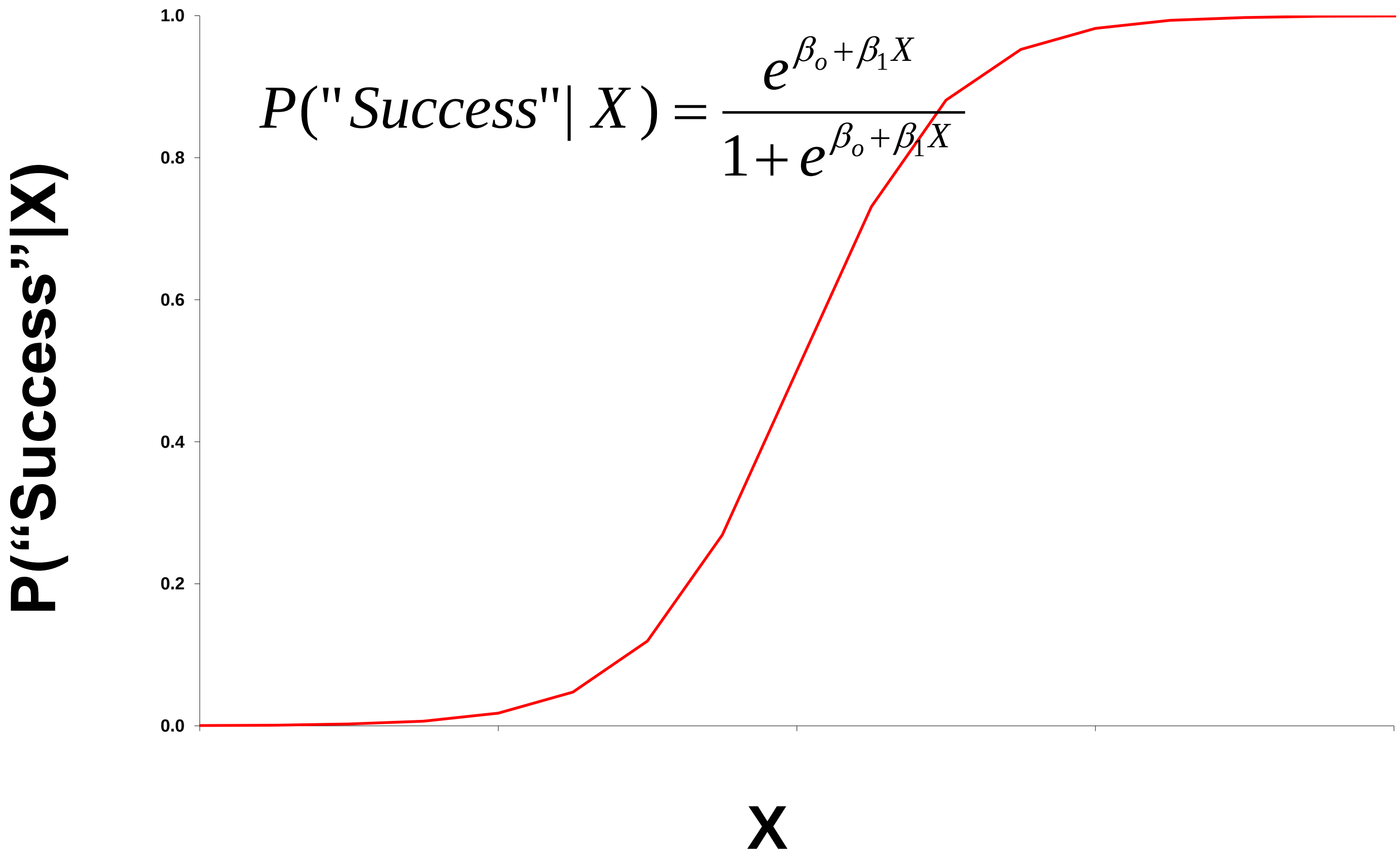
Logistic Curve



Sigmoid cure

Logistic Function

Logistic Function



$$P("Success"|X) = \frac{e^{\beta_o + \beta_1 X}}{1 + e^{\beta_o + \beta_1 X}}$$

Logit Transformation

- The logistic regression model is given by

$$P(Y \mid X) = \frac{e^{\beta_o + \beta_1 X}}{1 + e^{\beta_o + \beta_1 X}}$$

- which is equivalent to

$$\underbrace{\ln\left(\frac{P(Y \mid X)}{1 - P(Y \mid X)}\right)}_{\textit{This is called the Logit Transformation}} = \beta_o + \beta_1 X$$

This is called the
Logit Transformation

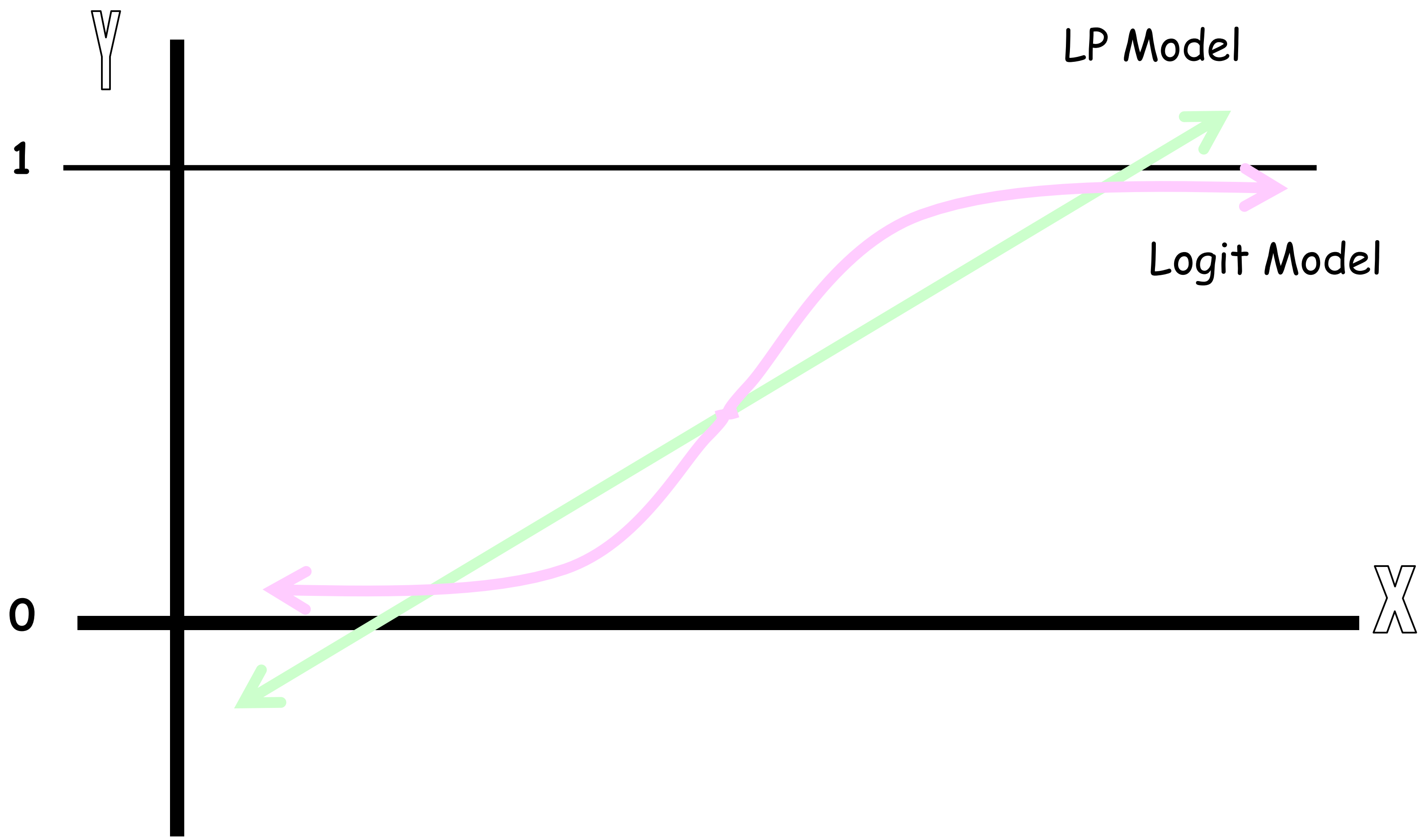
Logit Transformation

- Logistic regression models transform probabilities called *logits*.

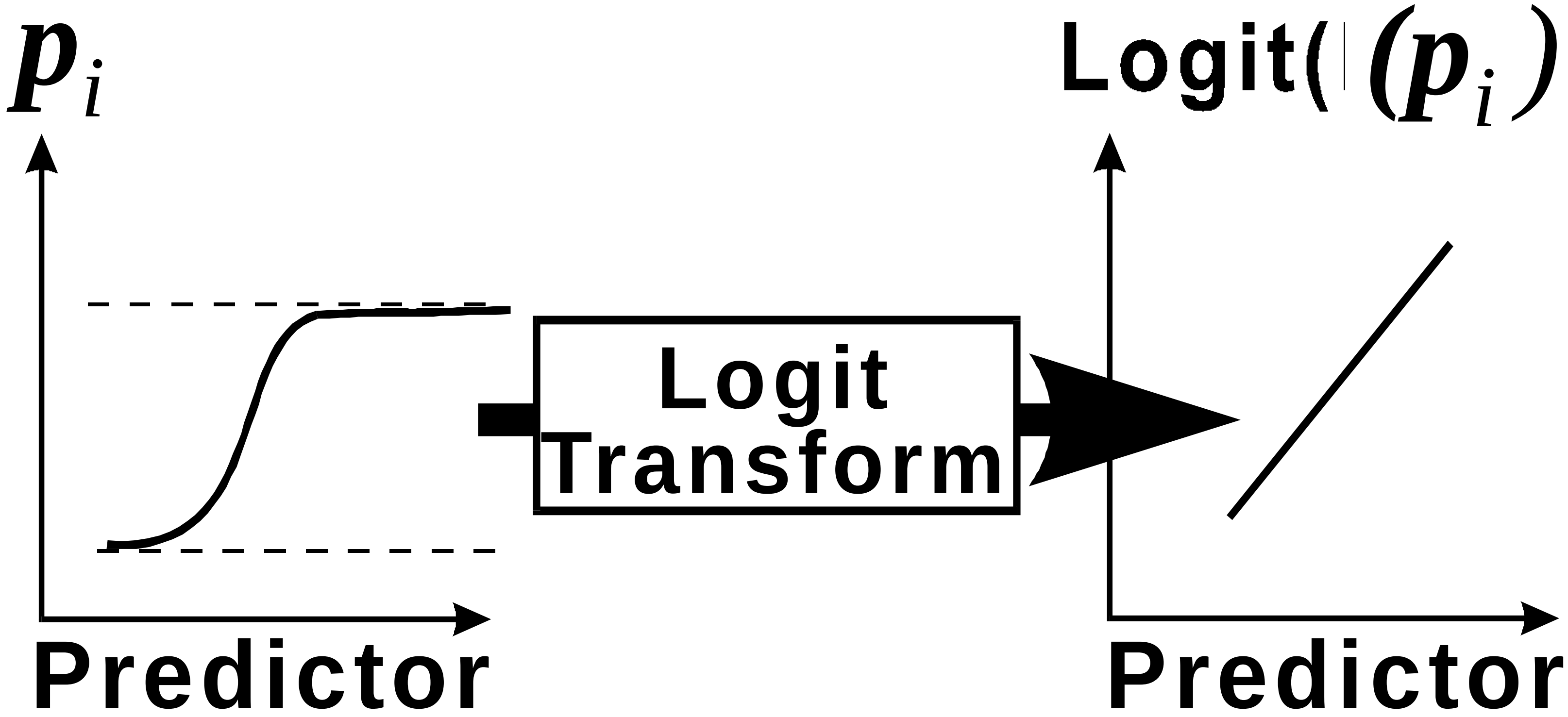
$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$$

- where
  - i indexes all cases (observations).
  - p_i is the probability the event (a sale, for example) occurs in the i^{th} case.
  - log is the natural log (to the base e).

Comparing LP and Logit Models



Assumption



Logistic regression model with a single continuous predictor

- $logit(p_i) = \log(odds) = \beta_0 + \beta_1 X_1$
- Where $logit(p_i)$ logit transformation of the probability of the event.
- β_0 intercept of the regression line
- β_1 slope of the regression line

The logistic Regression Model

- Let p denote $P[y = 1] = P[\text{Success}]$. This quantity will increase with the value of x.

The ratio: $\dfrac{p}{1-p}$ is called the **odds ratio**

This quantity will also increase with the value of x, ranging from zero to infinity.

The quantity: $\ln\left(\dfrac{p}{1-p}\right)$

is called the **log odds ratio**

Example: odds ratio, log odds ratio

Suppose a die is rolled:

Success = "roll a six", $p = 1/6$

The **odds ratio**

$$\frac{p}{1-p} = \frac{\frac{1}{6}}{1-\frac{1}{6}} = \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5}$$

The **log odds ratio**

$$\ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{1}{5}\right) = \ln(0.2) = -1.69044$$

The logistic Regression Model

Assumes the **log odds ratio** is linearly related to *x*.

i. e. : $$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

In terms of the **odds ratio**

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x}$$

The logistic Regression Model

Solving for p in terms x.

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x}$$

$$p = e^{\beta_0 + \beta_1 x}\left(1 - p\right)$$

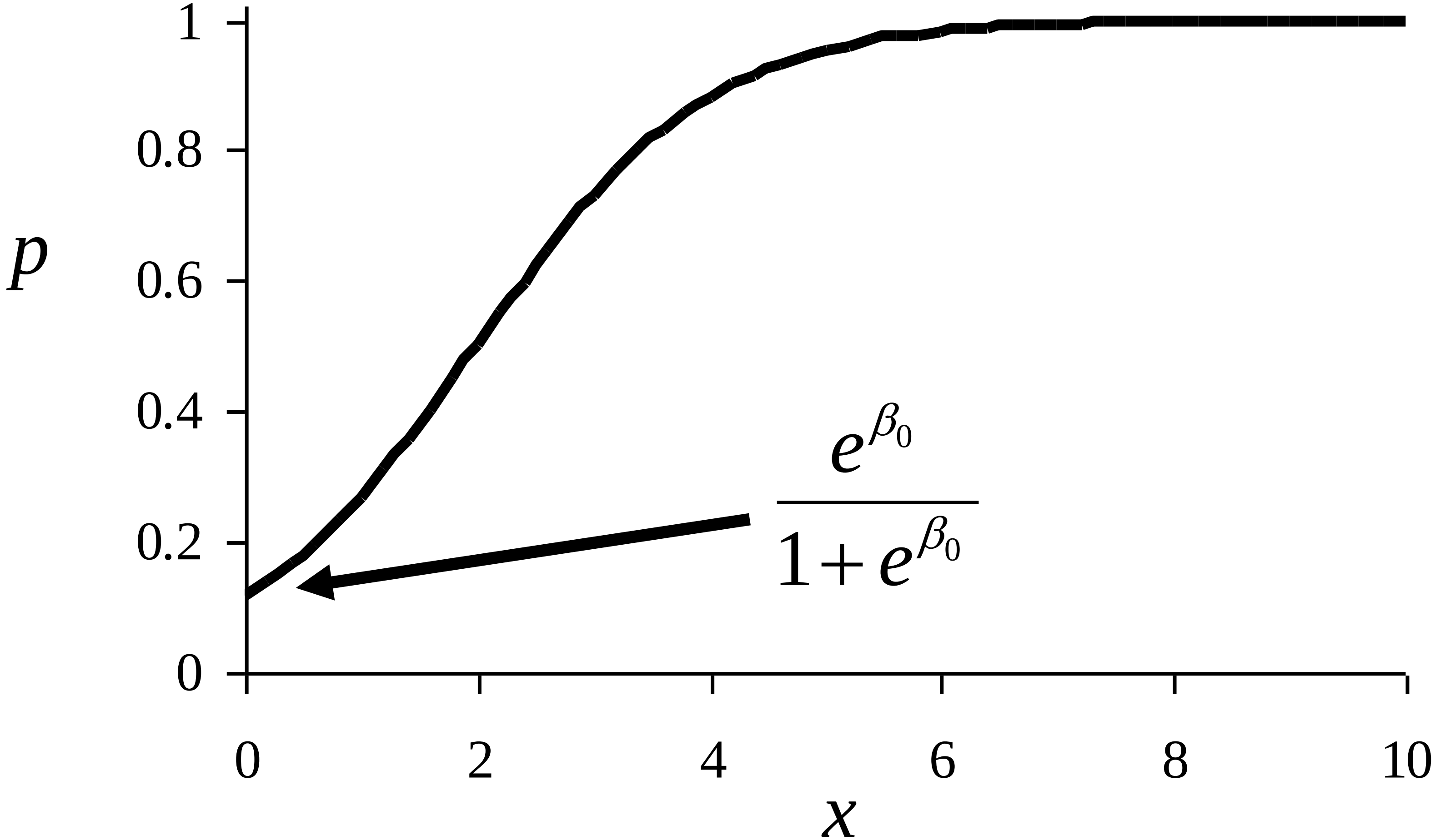
$$p + pe^{\beta_0 + \beta_1 x} = e^{\beta_0 + \beta_1 x}$$

or

$$\boxed{p = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}}$$

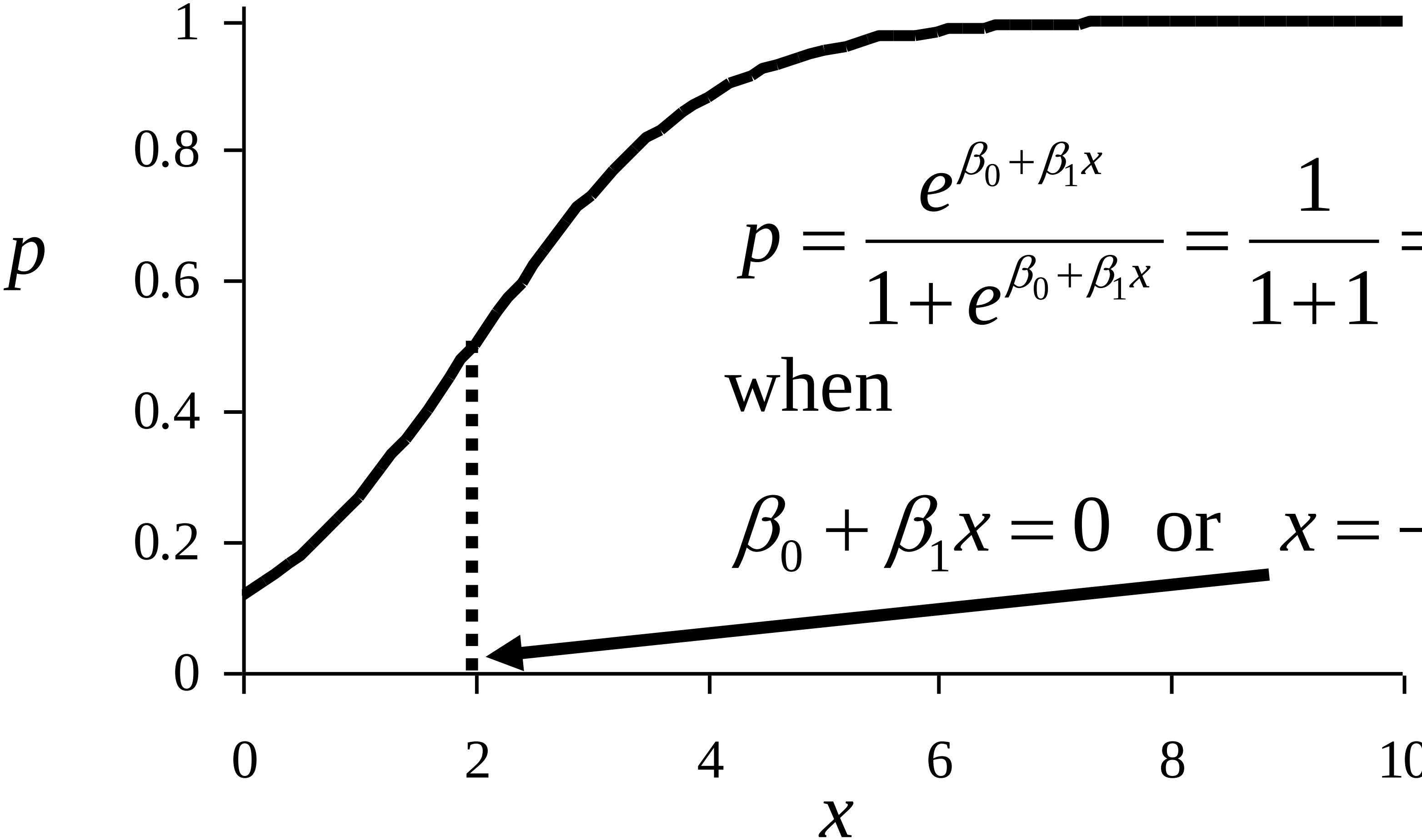
Interpretation of the parameter β_0

- determines the intercept



Interpretation of the parameter β_1

- determines when *p* is 0.50 (along with β_0)



$$p = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1+1} = \frac{1}{2}$$

when

$$\beta_0 + \beta_1 x = 0 \quad \text{or} \quad x = -\frac{\beta_0}{\beta_1}$$

Interpretation of the parameter β_1...

Also $$\frac{dp}{dx} = \frac{d}{dx}\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

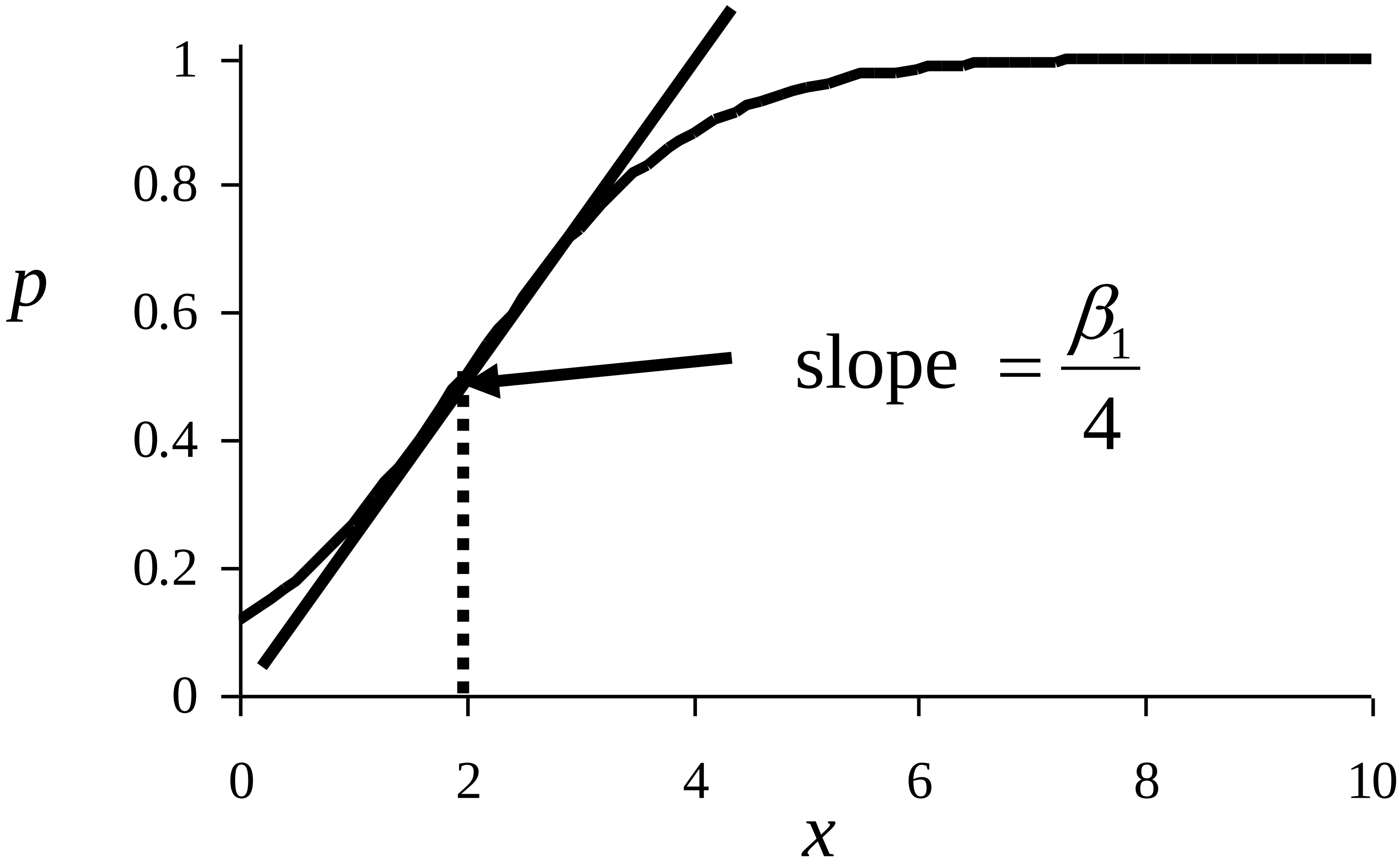
$$= \frac{e^{\beta_0+\beta_1 x}\beta_1\left(1+e^{\beta_0+\beta_1 x}\right) - e^{\beta_0+\beta_1 x}\beta_1 e^{\beta_0+\beta_1 x}}{\left(1+e^{\beta_0+\beta_1 x}\right)^2}$$

$$= \frac{e^{\beta_0+\beta_1 x}\beta_1}{\left(1+e^{\beta_0+\beta_1 x}\right)^2} = \frac{\beta_1}{4} \text{ when } \quad x = -\frac{\beta_0}{\beta_1}$$

$\frac{\beta_1}{4}$ is the rate of increase in p with respect to x when $p = 0.50$

Interpretation of the parameter β_1

determines slope when *p* is 0.50



Binary Classification

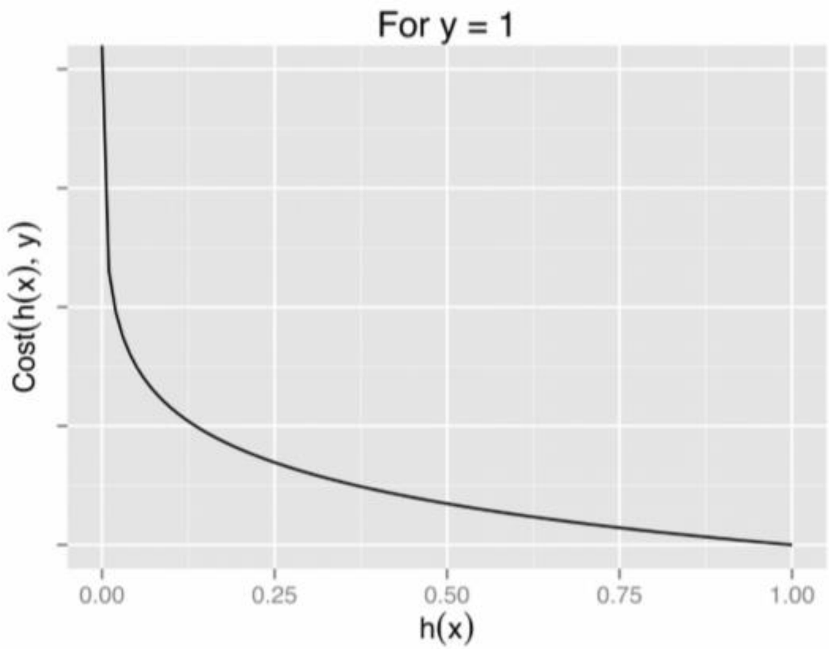
- In logistic regression we take two steps:
  - First step yields estimates of the probabilities of belonging to each class. In the binary case we get an estimate of $P(Y = 1)$.
  - the probability of belonging to class 1 (which also tells us the probability of belonging to class 0).
- In the next step we use
  - a cutoff value on these probabilities in order to classify each case to one of the classes.
  - a cutoff of 0.5 means that cases with an estimated probability of $P(Y = 1) > 0.5$ are classified as belonging to class 1,
  - whereas cases with $P(Y = 1) < 0.5$ are classified as belonging to class 0.
  - The cutoff need not be set at 0.5.

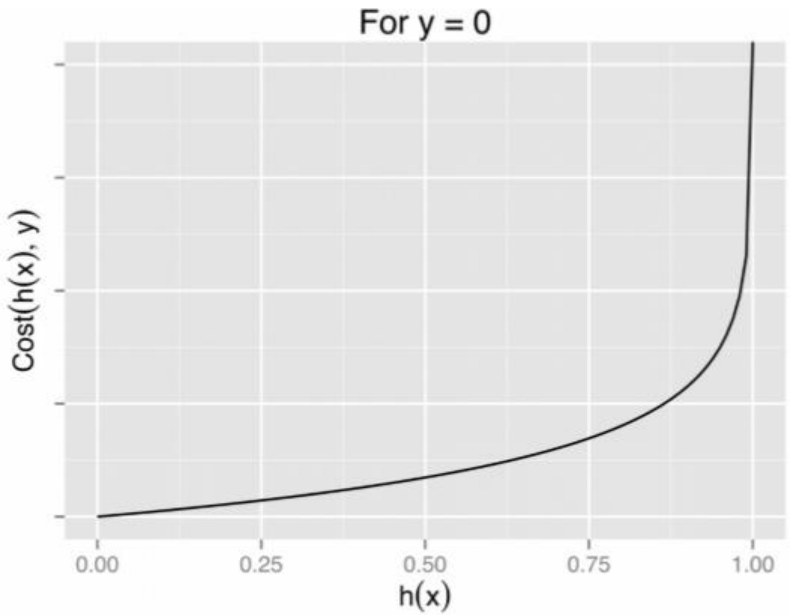
Types of Logistic Regression

- Binary Logistic Regression
  - The categorical response has only two 2 possible outcomes. Example: Spam or Not
- Multinomial Logistic Regression
  - Three or more categories without ordering. Example: Predicting which food is preferred more (Veg, Non-Veg, Vegan)
- Ordinal Logistic Regression
  - Three or more categories with ordering. Example: Movie rating from 1 to 5

Cost Function

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





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

Gradient Descent

- Now the question arises, how do we reduce the cost value. Well, this can be done by using Gradient Descent.
- The main goal of Gradient descent is to minimize the cost value. i.e. min $J(\theta)$.
- Now to minimize our cost function we need to run the gradient descent function on each parameter i.e.

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

Gradient Descent…

- Objective: To minimize the cost function we have to run the gradient descent function on each parameter

Want $\min_\theta J(\theta)$:

Repeat $\{$

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

(simultaneously update all θ_j)

$\}$