Machine Learning with C/C++

WELCOME

Day 2: Logistic Regression

Today’s Agenda

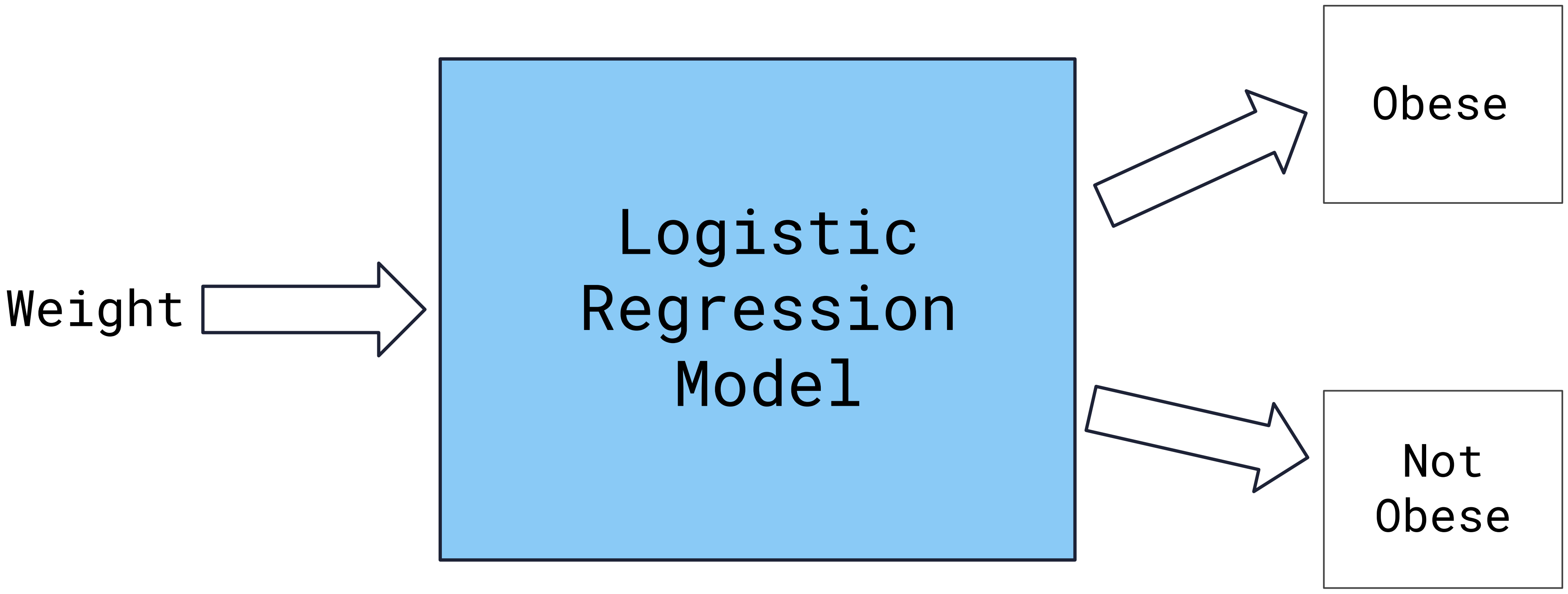
0. Today’s Agenda(You’re here)
1. Logistic Regression
2. Why not just use Linear Regression?
3. Hypothesis Function
4. Cost/Error Function
5. Gradient Descent
6. Logistic Regression in C

Logistic Regression

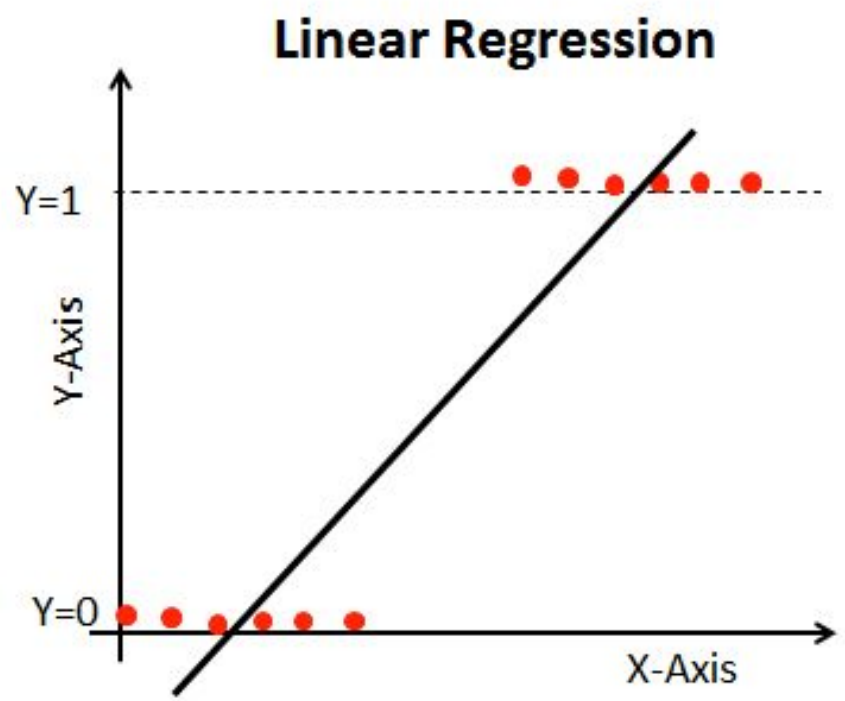
Definition

Draw the graph

Logistic Regression, in the most general sense, is the process of classifying input data into known categories.



Why not just use linear regression?


Linear Regression
Y=1
Y-Axis
Y=0
X-Axis

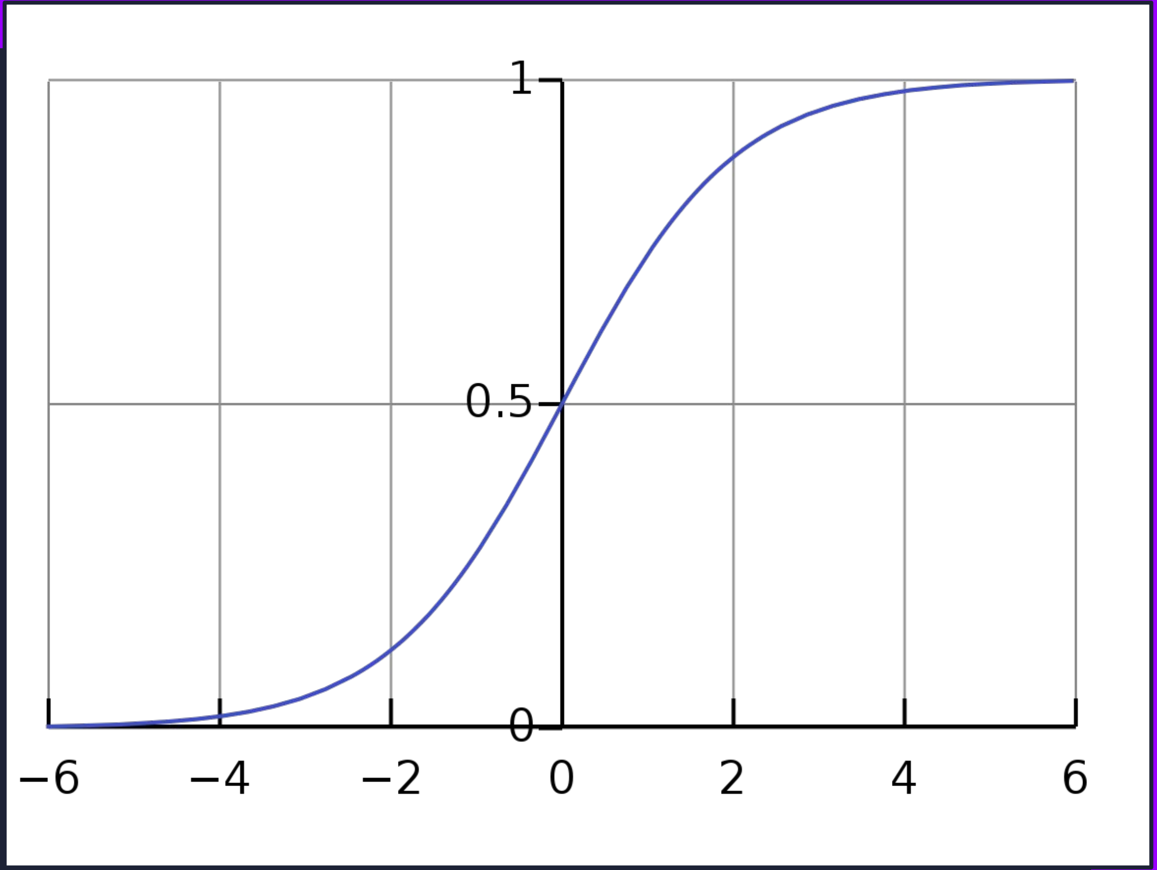
01

Hypothesis

As discussed, the equation that we require to best fit to the data is called the hypothesis function. For logistic regression, it is generally written as,

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

Sigmoid Curve



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

The Sigmoid Function takes in all values from $(-\infty,+\infty)$ and maps them to a unique value between
(0, 1).

Error/Cost Function

2.0

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

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x), y)$$

We could plug our present hypothesis function (the sigmoid function) but that makes the cost function graph non-convex.

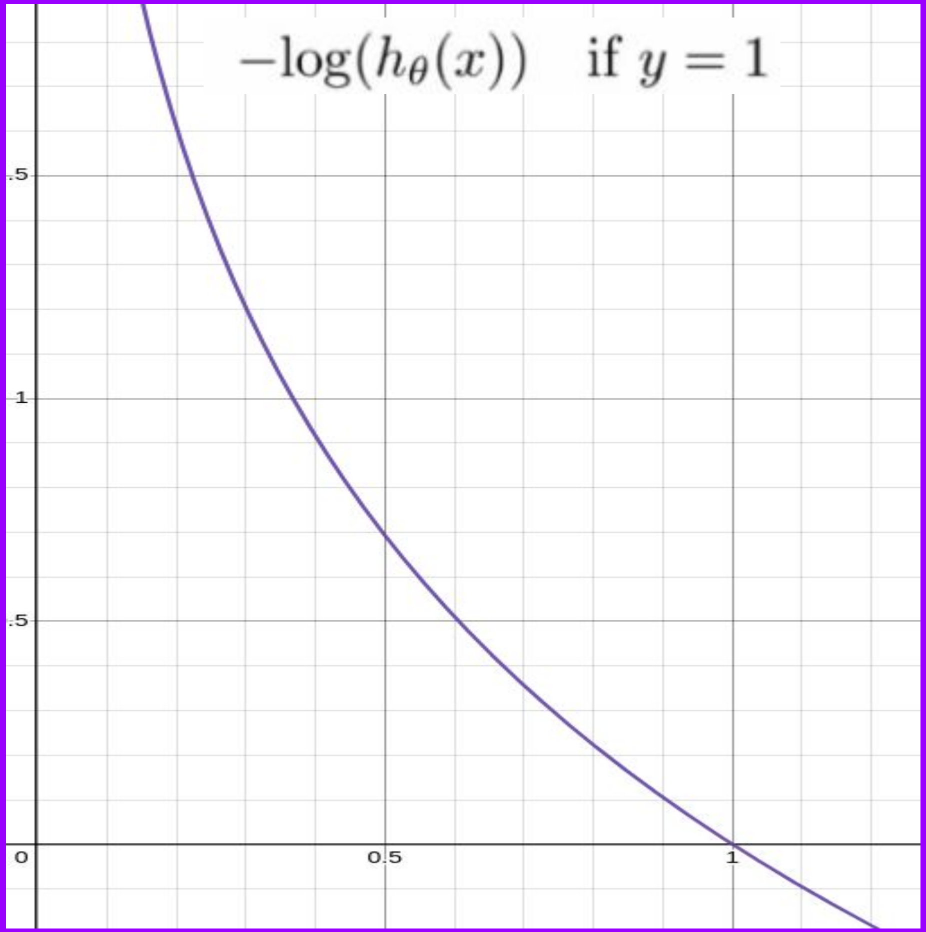
Cost Function for Logistic Regression

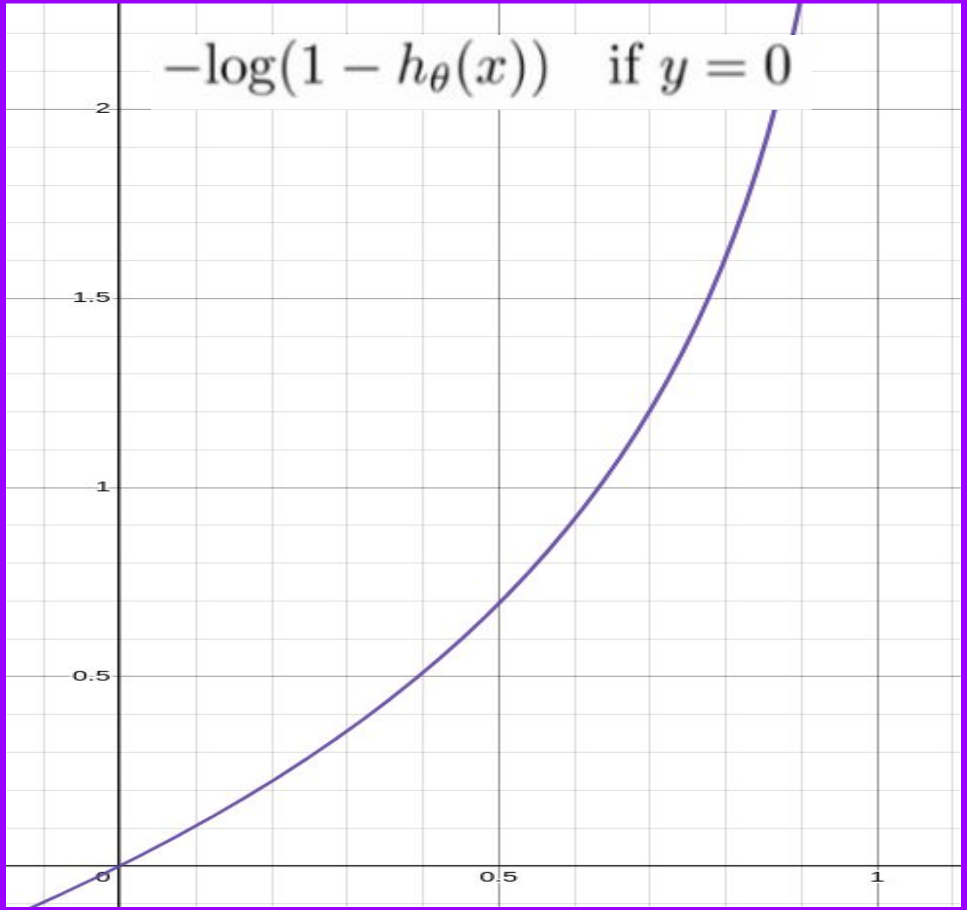
2.1

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x), y)$$

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

Graphs of cost function





2.2 Simplified Cost Function

$$Cost(h_\theta(x), y) = -ylog(h_\theta(x)) - (1-y)log(1-h_\theta(x))$$

2.3 Final Error Function

$$J(\theta) = -\frac{1}{n} \cdot \sum_{i=0}^{n} [ylog(h_\theta(x)) + (1-y)log(1-h_\theta(x))]$$

3. GRADIENT DESCENT

Gradient Descent is an iterative algorithmic technique to find the minimum of a general nth dimensional function. It is expressed as:

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

Applying it for the Simplified Cost Function of Logistic Regression:

$$\frac{\partial}{\partial x}\left(-\frac{1}{n} \cdot \sum_{i=0}^{n} y log(h_\theta(x)) + (1-y) log(1-h_\theta(x))\right)$$

Solving the partial derivative

$$\frac{\partial}{\partial\theta_j}(J(\theta)) = \frac{1}{m}\sum_{i=0}^{n}(h_\theta(x^{(i)}) - y^{(i)}) \cdot x_{(j)}^{(i)}$$

It looks familiar, right?

And so the Finally, the gradient descent algorithm is the same as for linear regression:

Repeat until convergence

{

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

}

(Simultaneously update all parameters)

“Talk is cheap, show me the code”

– Linus Torvalds

xkcd.com/303


THE #1 DATA SCIENTIST EXCUSE FOR LEGITIMATELY SLACKING OFF:
"MY MODEL'S TRAINING"
HEY! GET BACK TO WORK!
TRAINING!
OH. CARRY ON.

Thank you!

Do you have any questions?

jenishpanta@gmail.com
9841551131(Not guaranteed to pick up)