Lecture 4

INTRODUCTION AND APPLICATIONS OF MACHINE LEARNING

Contacts

Instructor

- Shreyas Bhat - 9082080984
- S I Harini - 7021247073
- Rishav Mukherji - 9920741703

Mentor

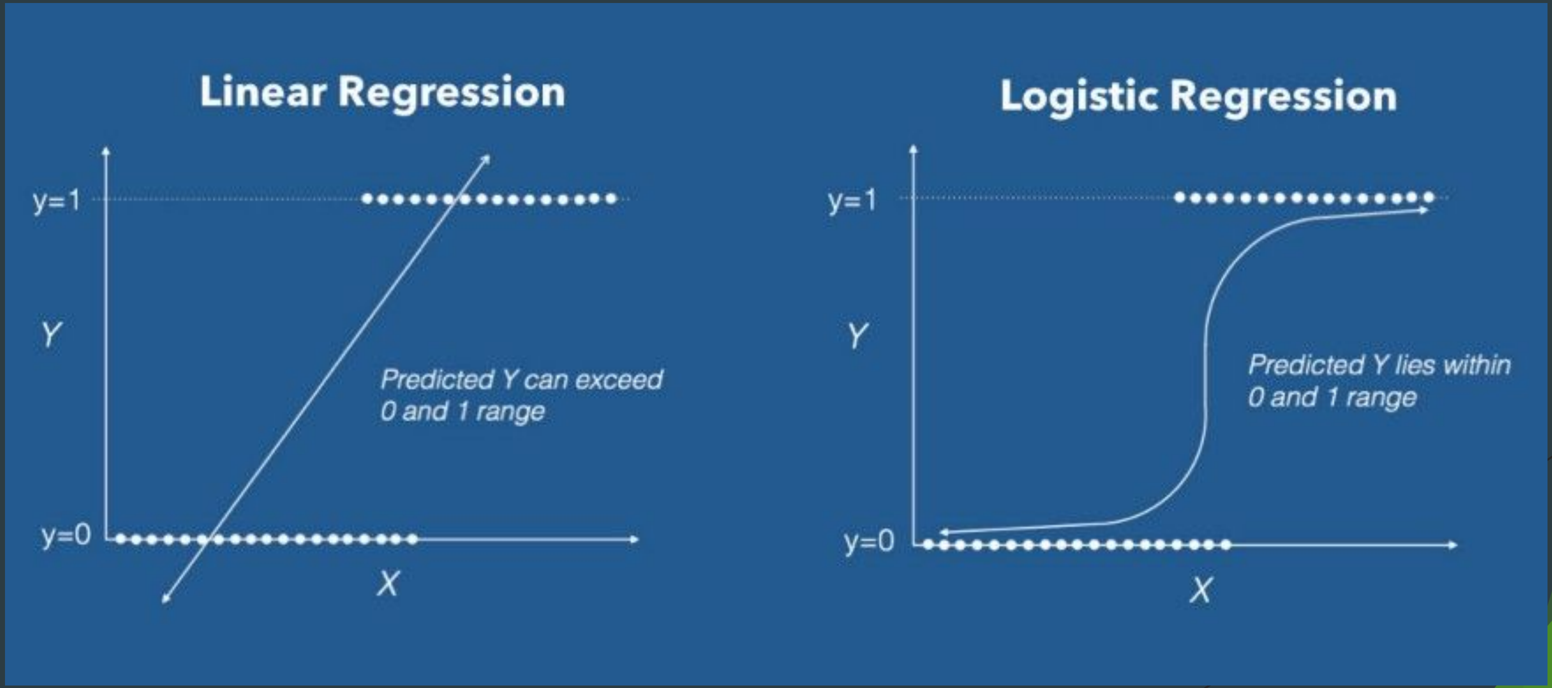
- Hardik Shah - 9427925103

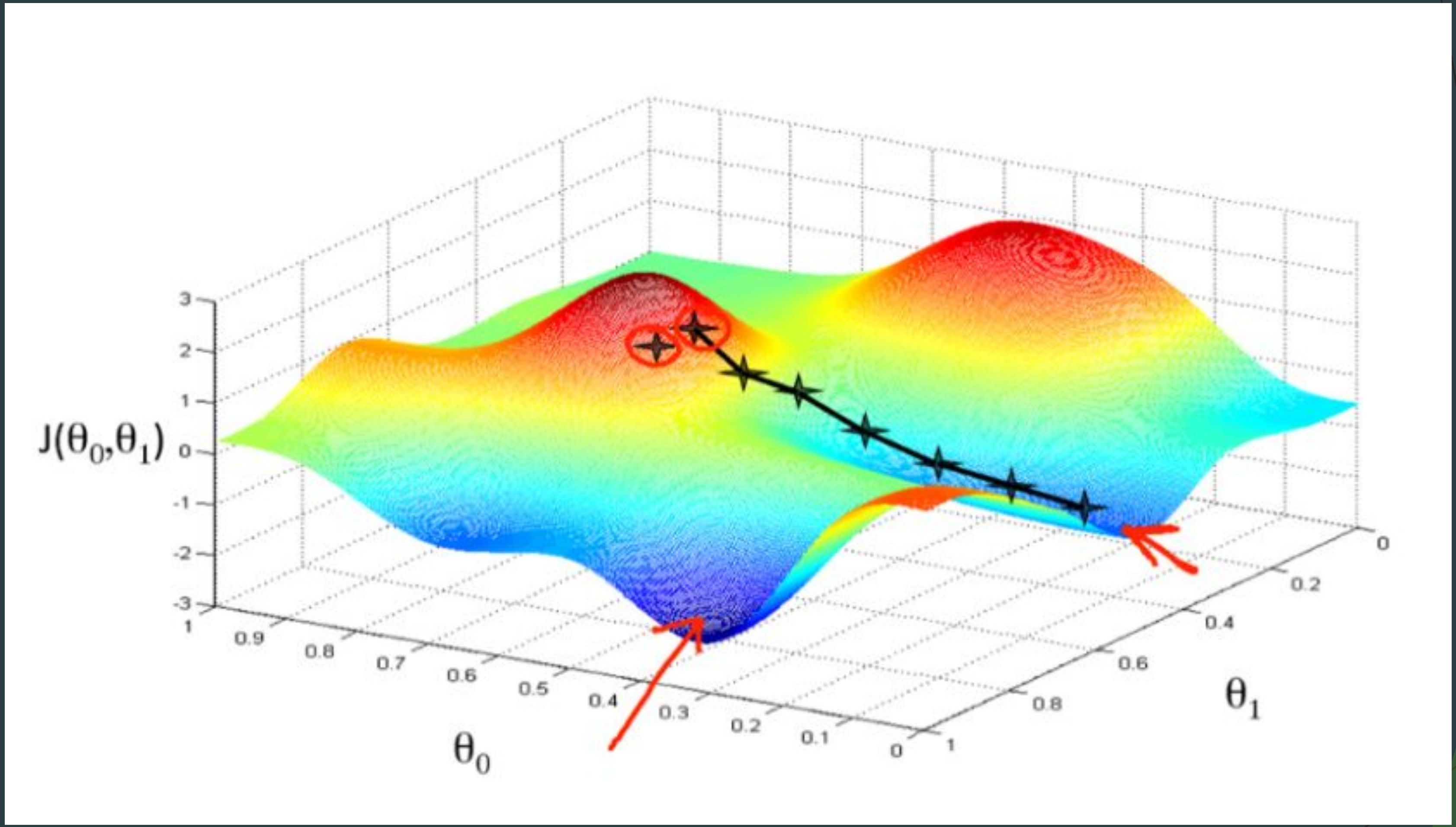
Roadmap

- Logistic regression
  - Hypothesis function
  - Loss function
  - Gradient descent update

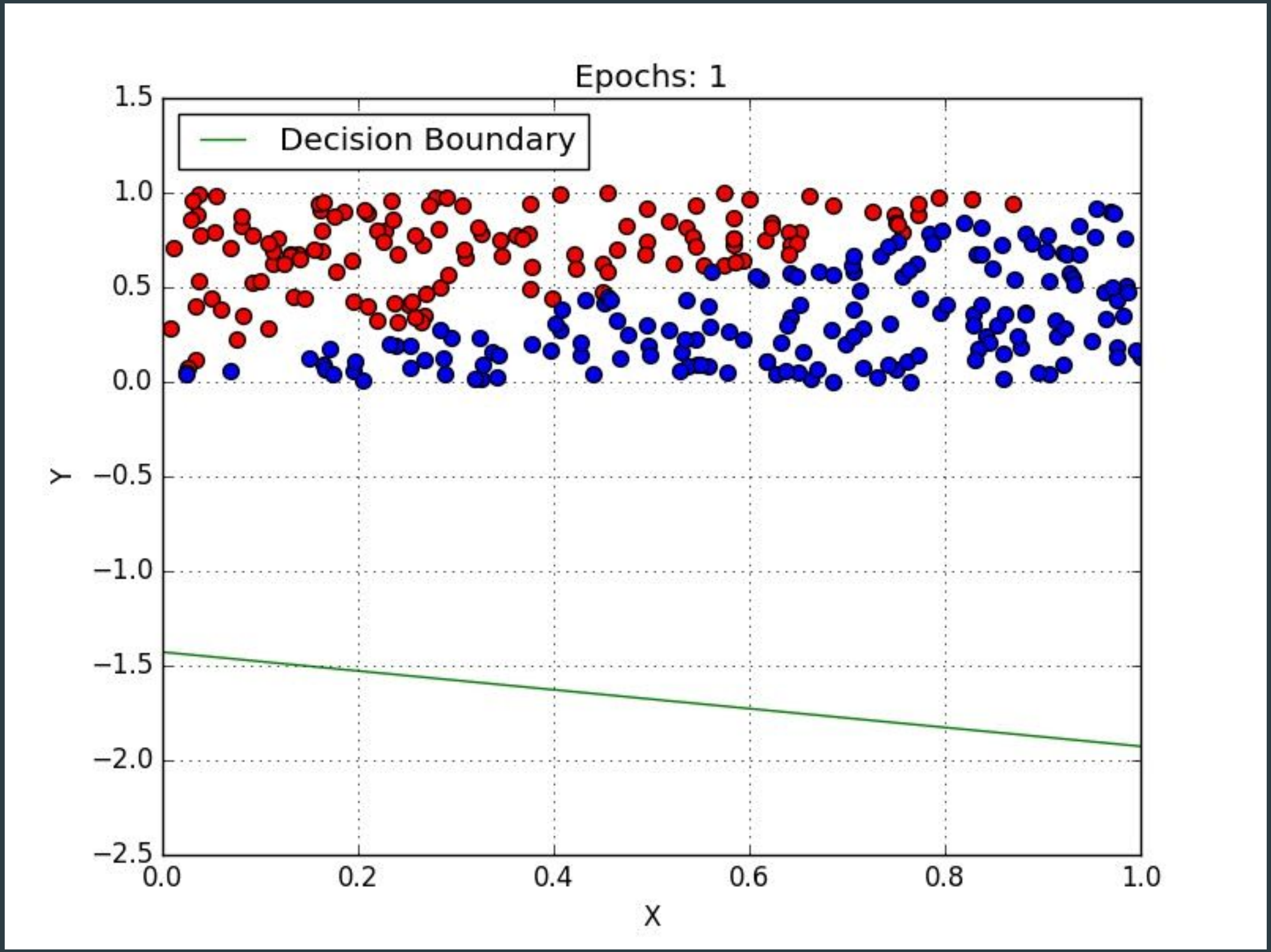
Logistic regression

- Logistic regression is a classification algorithm used to assign observations to a discrete set of classes. Some of the examples of classification problems are Email spam or not spam, Online transactions Fraud or not Fraud, Tumor Malignant or Benign. Logistic regression transforms its output using the logistic sigmoid function to return a probability value.



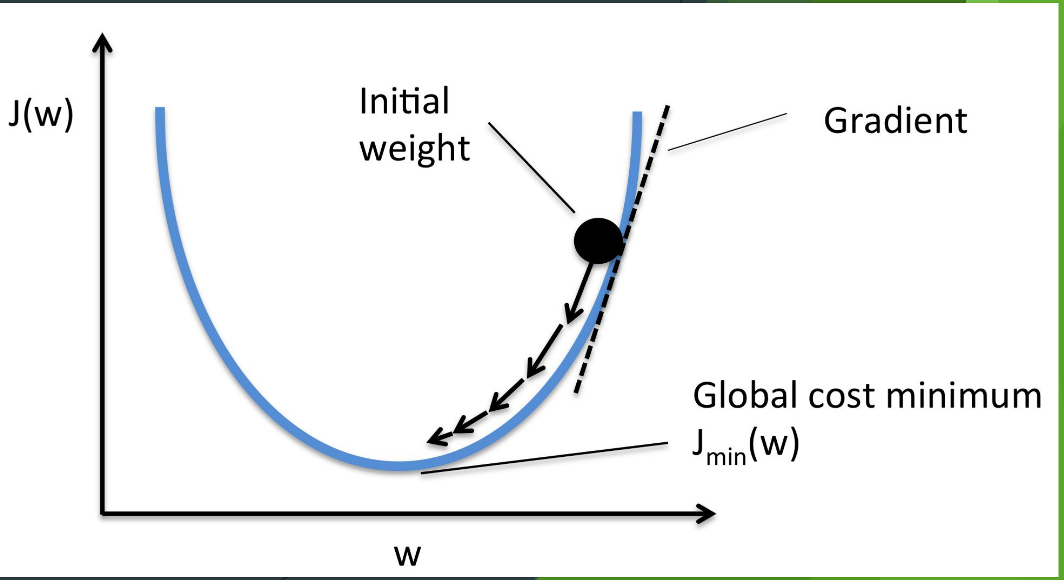


J(θ0,θ1)
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ0
0
0.2
0.4
0.6
0.8
1
θ1

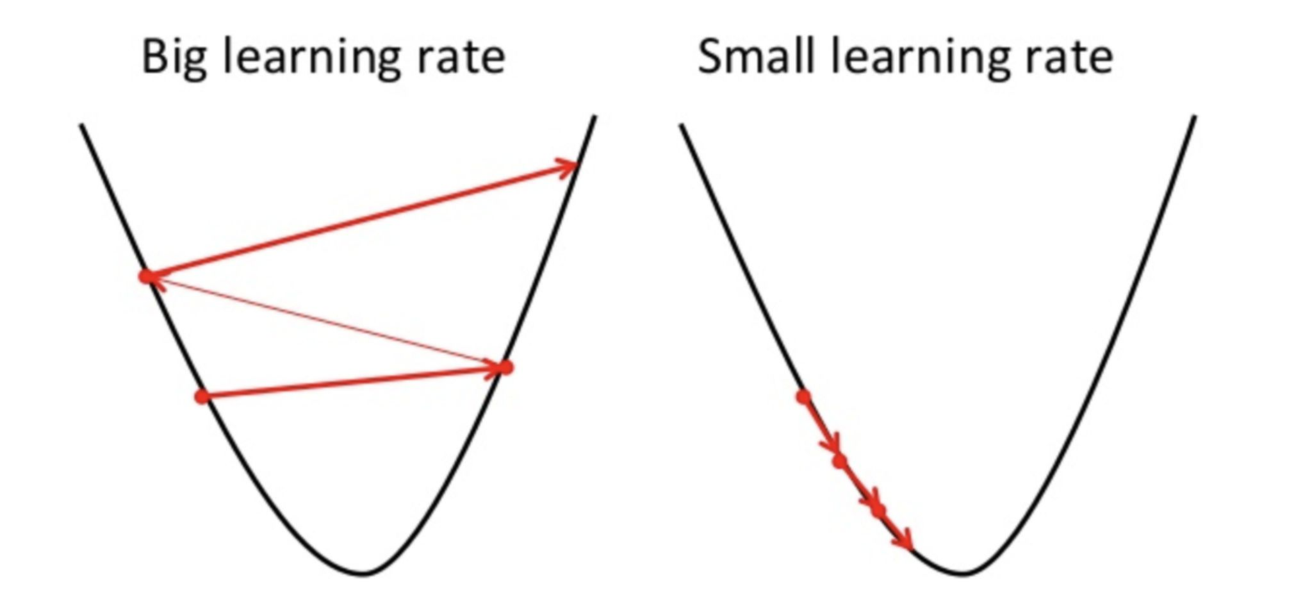

Epochs: 1
Decision Boundary
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
Y
0.0
0.2
0.4
0.6
0.8
1.0
X

Gradient Descent

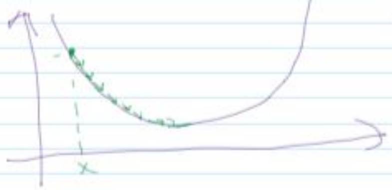
- Optimization method to find minima in a function (here reduce the cost function MSE)
- It is an iterative process
- Start at any point and move towards the minima
- This depends on
  - Step size (η)
  - direction(determined by the negative of the gradient)



Gradient update



gradient descent algorithm



$y = f(x)$

$f(x) = (x-5)^2$

constant (learning rate)

$$X = X - \eta \underbrace{\left[\frac{\partial f(x)}{\partial x}\right]}_{\text{gradient/slope}} \to 0$$

1 θ init

2 $J_\theta = \frac{1}{m} \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)}_{actual}\right]^2$

3 update θ using GD

$J(\theta) = \frac{1}{m} \sum \left[\theta_0 + \theta_1 x\right]$

$$\theta = \theta - \eta \nabla J(\theta) \to \theta_0, \theta_1$$

$$\Rightarrow \quad \theta_0 = \theta_0 - \eta \frac{\partial J(\theta)}{\partial \theta_0} \qquad \theta_1 = \theta_1 - \eta \frac{\partial J(\theta)}{\partial \theta_1}$$

$$J(\theta) = \frac{1}{m} \sum_{i=0}^{m} \left[\theta_0 + \theta_1 x^{(i)} - y^{(i)}_{actual}\right]^2$$

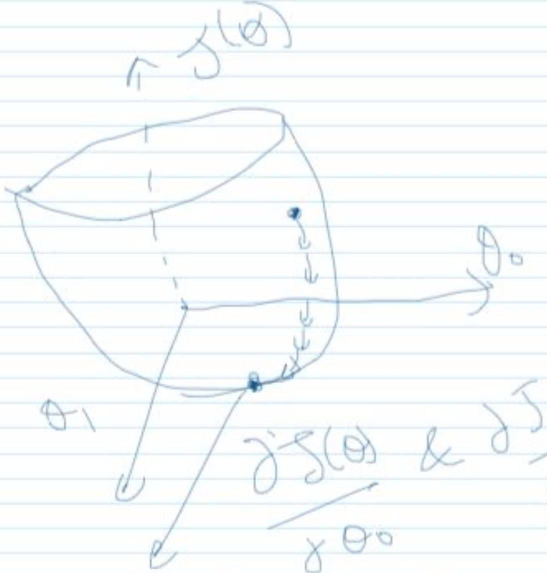
① $$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{2}{m} \sum_{i=0}^{m} \left[\theta_0 + \theta_1 x^{(i)} - y^{(i)}_{actual}\right] \cdot 1$$

② $$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{2}{m} \sum_{i=0}^{m} \left[\theta_0 + \theta_1 x^{(i)} - y^{(i)}_{actual}\right] \cdot x^{(i)}$$

finally,

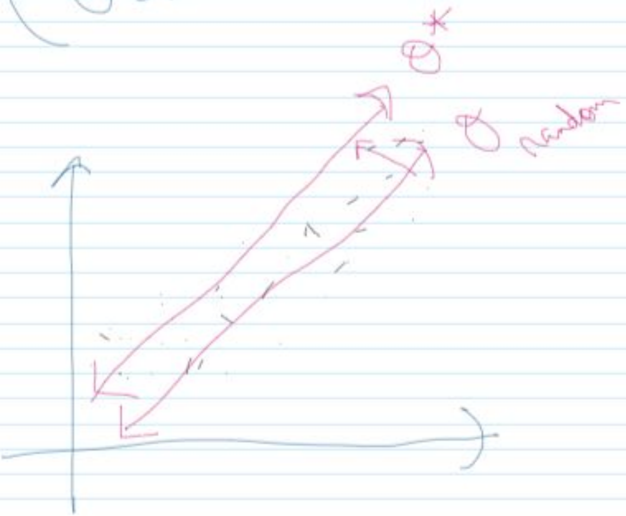
$$\theta_0 = \theta_0 - \eta \frac{1}{m} \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)}_{actual}\right] 1$$

$$\theta_1 = \theta_1 - \eta \frac{1}{m} \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)}_{actual}\right] x^{(i)}$$



$$\frac{\partial J(\theta)}{\partial \theta_0} \ \& \ \frac{\partial J(\theta)}{\partial \theta_1} =$$

$$(\theta_0, \theta_1) \to \theta^*$$



21,00,000
2,00,000

* Bedrooms	* Area	* Location	₹
3	2	300	100000
4	2	400	150000

$$X = \begin{bmatrix} x_0^1 & x_1^1 & \cdots & x_n^1 \\ & \vdots & & \\ \text{——} & x_j^i & \text{——} & \\ & \vdots & & \end{bmatrix} = \begin{bmatrix} \text{——} \ x^{(1)} \ \text{——} \\ \text{——} \ x^{(2)} \ \text{——} \\ \vdots \\ \text{——} \ x^{(m)} \ \text{——} \end{bmatrix}$$

(m examples × n features)

$i \rightarrow$

$x \quad j \rightarrow$

area, loc

$(x_1, x_2, x_3 \ldots)$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \cdots \theta_n x_n$$

$$= \theta_0 + \sum_{i=1}^{n} \theta_i x_i \qquad x_0 = 1$$

$$= \theta_0 x_0$$

$$= \sum_{i=0}^{n} \theta_i x_i$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$X = \begin{array}{c} \begin{array}{cccc} x_0 & x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 1 & \text{————} \\ 1 & \text{————} \\ 1 & \text{————} \\ 1 & \text{————} \end{bmatrix}_{m \times (n+1)} \end{array}$$

n+1 parameters

$\theta_0 - \theta_n$

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta)$$

$$J(\theta) = \frac{1}{m} \sum \left(\hat{y} - y^{(i)}\right)^2$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \left(h_\theta(x) - y\right)^2$$

$$= 2\left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(h_\theta(x)\right)$$

$$= \left(h_\theta(x) - y\right) \frac{\partial}{\partial \theta_j} \left[\sum_{j=0}^{n} \theta_j x_j\right]$$

$$= \left(h_\theta(x) - y\right) \cdot x_j$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} \left(\hat{y} - y\right) \cdot x_j$$

gradient w.r.t

θ_j

$j = 0$

$j = 1$

$$\boxed{\frac{\partial J(\theta)}{\partial \theta_0} = \sum_{i=1}^{m} \left(\hat{y} - y\right) \cdot 1 \qquad \frac{\partial J(\theta)}{\partial \theta_1} = \sum_{i=1}^{m} \left(\hat{y} - y\right) \cdot x_i}$$

Gradient update

$$\theta_j = \theta_j - \eta \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) x_j$$

Logistic regression

→ supervised algorithm (X, y)

→ Classification algorithm

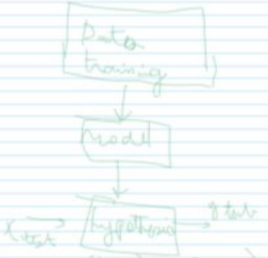
Binary

DT SVM

Spam 0

not spam 1

$X \in R^n \longrightarrow y \in \{0, 1\}$



$g(h(x)) \in (0, 1)$

$g(z) = 0 \quad z \le 0$

$g(z) = 1 \quad z > 0$

$g(z) = \frac{1}{1+e^{-z}}$

$g(z) \geq 0.5 \Rightarrow$ class 1

$g(z) < 0.5 \Rightarrow$ class 0

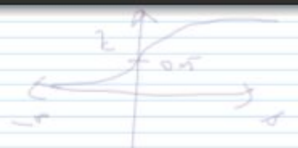
$g(z) \to 1 \quad z \to \infty$

$g(z) \to 0 \quad z \to -\infty$

$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i$

(1) $g(h_\theta(x)) = \frac{1}{\ldots - \sum \theta_i x_i}$

Cat = 1 dog = 0



$g(z) \geq 0.5 \Rightarrow$ class 1

$g(z) < 0.5 \Rightarrow$ class 0

$$g(z) = \frac{1}{1+e^{-z}}$$

$g(z) \to 1 \quad z \to \infty$

$g(z) \to 0 \quad z \to -\infty$

$x_{test} \to$ hypothesis

$g(h(x)) \in (0, 1)$

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i$$

(1) $$g(h_\theta(x)) = \frac{1}{1 + e^{-\sum_{i=0}^{n} \theta_i x_i}}$$

(1) (0, 1)

(2) Assign diff value for diff classes → classifying

(2) Loss func

$$\text{Avg. Sq loss} = \frac{1}{m} \sum \left(\hat{y}^{(i)} - g(h_\theta(x^{(i)}))\right)^2$$

X

Cat = 1 dog = 0

$(1 - 0.8)^2$

$+ (1 - 0.2)^2$

$+ \vdots$

$(0 - 0.6)^2$

total error

(3) Minimize the error

$$\theta = \theta - \eta \frac{\partial L}{\partial \theta}$$

Log loss Binary Cross entropy $\hat{y} = h_\theta(x)$

$$\boxed{\text{loss} = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log g(h_\theta(x^{(i)})) + \underbrace{(1 - y^{(i)})\left(1 - \log g(h_\theta(x^{(i)}))\right)}_{0}}$$

$y = \{1\ 0\ 1\ 1\ 0\}$

$= \{0.5\ 0.6\ 0.2\ 0.1\ 0.3\}$

$y^{(i)} = 0\ 0\ 0\ 0\ 0$

$$y = \{1\ 0\ 1\ 1\ 0\} = \{0.5\ 0.6\ 0.7\ 0.1\ 0.3\}$$

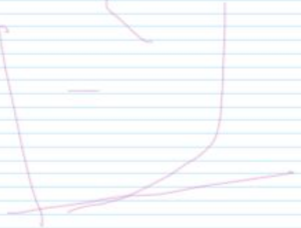
$$y^{(i)} = 1\ 1\ 1\ 1\ 1$$

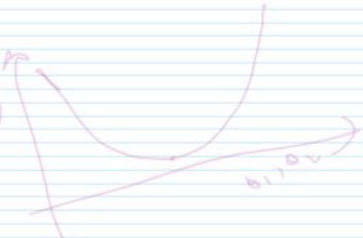
$$loss = -\frac{1}{m}\sum_{i=1}^{m} \log\left(h_\theta(x^{(i)})\right)$$

$+$

$$y^{(i)} = 0\ 0\ 0\ 0\ 0$$

$$loss = -\log\left(1 - h_\theta(x^{(i)})\right)$$





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

$$\boxed{h_\theta(x) = g\left(\Sigma x_i \theta_i\right)}$$

$$\Rightarrow \theta = \theta - \eta \cdot \frac{\partial J(\theta)}{\partial \theta}$$

THANK YOU!