DLSG Week 2: Logistic Regression

Thuong Nguyen Canh

July 2019

Binary Classification

- Learn a function to classify a given set $\boldsymbol{x}$ into two groups (y)

Is it a dog?

Dog / Non Dog

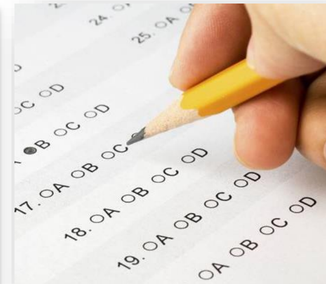
Drug test

Effective/Not-Effective

Medical Test

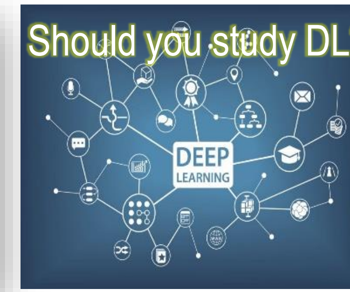
Positive/Negative

Will you pass?



Pass/ Fail

Decision Making



Yes/ No

- Notations

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} \rightarrow \boxed{\mathcal{F}(\cdot)} \rightarrow y$$

$\boldsymbol{x} \in \mathbb{R}^{n_x}$ $\quad$ $y \in \{0, 1\}$

Training pairs $\{(\boldsymbol{x}^{(1)}, y^{(1)}), \dots, (\boldsymbol{x}^{(m)}, y^{(m)})\}$

In matrix form

$$\boldsymbol{X} = \begin{bmatrix} | & | & & | \\ \boldsymbol{x}^{(1)} & \boldsymbol{x}^{(2)} & \cdots & \boldsymbol{x}^{(m)} \\ | & | & & | \end{bmatrix} \qquad \boldsymbol{y} = [y^{(1)}, \dots, y^{(m)}]$$

Inputs $\qquad$ Labels

(Non) Linear Classifier

- Learn a function to classify a given set $\boldsymbol{x}$ into two groups (y)

[32x32x3]

$$\mathcal{F}(\boldsymbol{x}, \boldsymbol{\omega}, b) = \sigma(\boldsymbol{\omega}^T \boldsymbol{x} + b) \longrightarrow$$

A binary number
Indicate yes or no

Parameters $\boldsymbol{\omega} \in \mathbb{R}^{n_x}, b \in \mathbb{R}^1$
Non Linear Activation $\boldsymbol{\sigma}(\cdot)$

- Example

5 pixels, 2 class

$$\underbrace{\begin{bmatrix} 0.3 & -0.7 & 0.1 & 2.3 & -1.0 \end{bmatrix}}_{\boldsymbol{\omega}} \underbrace{\begin{bmatrix} 30 \\ 89 \\ 226 \\ 180 \\ 28 \end{bmatrix}}_{\boldsymbol{x}^{(i)}} + \underbrace{\boxed{3.1}}_{b} \longrightarrow \boxed{358.4}$$

Yes

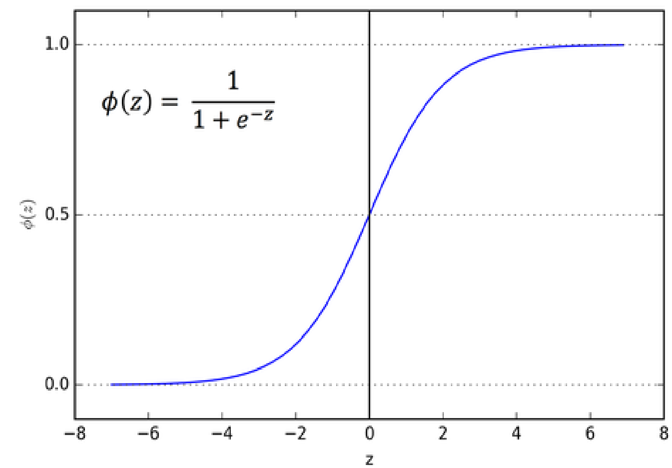
BxCx**WxH**
16x3x512x1024

Logistic Regression

- A non-linear regression with sigmoid activation
  - Given $\boldsymbol{x}$, estimate $\hat{y} = P(y = 1|\boldsymbol{x})$ as

$$\hat{y} = \boldsymbol{\sigma}(\boldsymbol{\omega}^T \boldsymbol{x} + b)$$

  - (Learnable) Parameters $\boldsymbol{\omega} \in \mathbb{R}^{n_x}, b \in \mathbb{R}^1$
  - Activation function $\boldsymbol{\sigma}$ (often be sigmoid function)



- Cost function:
  - How good is your prediction

Negative Loglikelihood

$$\mathcal{L}(\hat{y}, y) = -(y \log \hat{y} + (1-y)\log(1-\hat{y}))$$

- y = 1, → $\mathcal{L}(\hat{y}, y) = -\log \hat{y}$ → want large ŷ
- y = 0, → $\mathcal{L}(\hat{y}, y) = -\log(1-\hat{y})$ → want small ŷ

Square Error

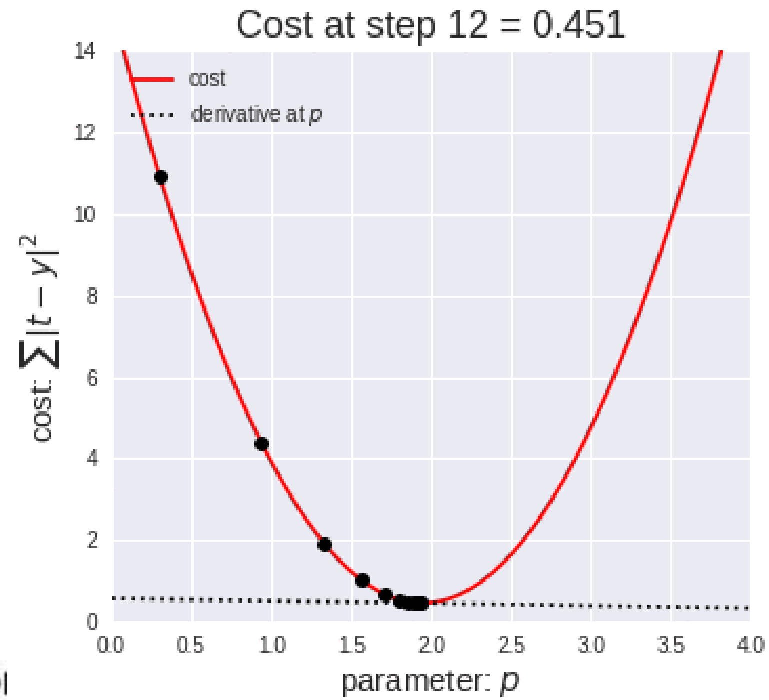
$$\cancel{\mathcal{L}(\hat{y}, y) = \frac{1}{2}(y-\hat{y})^2}$$

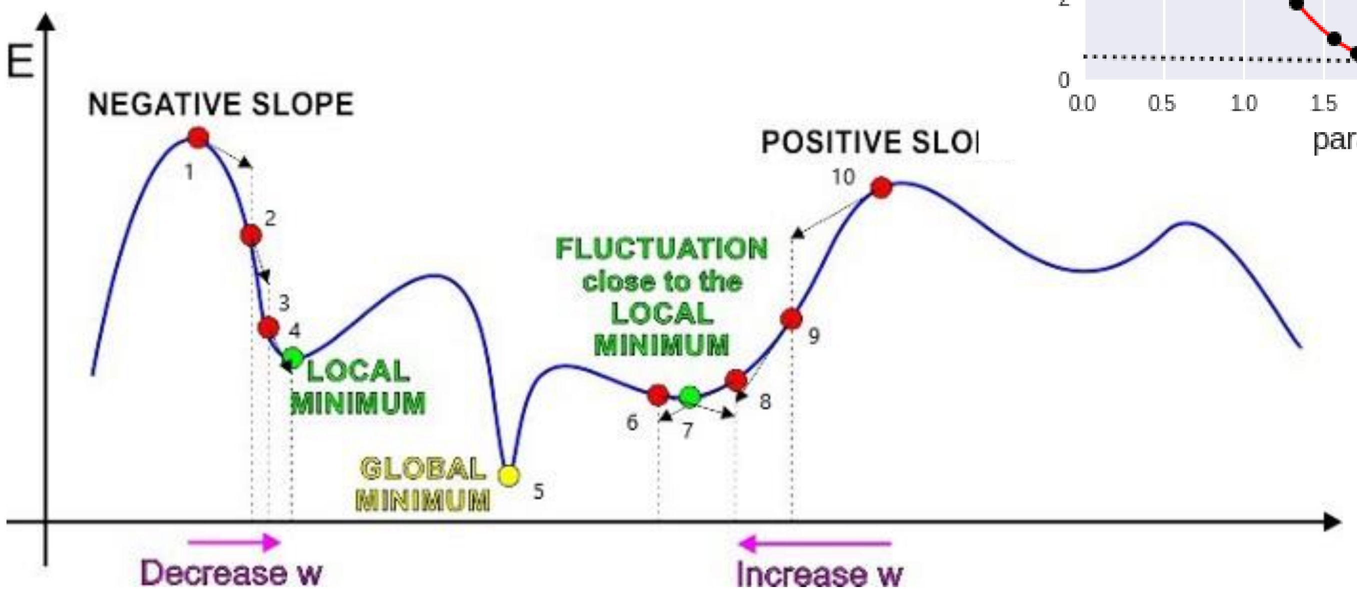
$$J(\omega, b) = \frac{1}{m}\sum \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$$

Gradient Descent

- Find ω, b to minimize loss function
- Optimization with gradient descent
  - Start at a "starting point"
  - Go quickly down as possible

Gradient – "Slop", Descent – "To go down"





Derivative

- Change rate of a function w.r.t. a variable at (near) a particular point

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$f(x,y) = xy \quad \rightarrow \quad \frac{\partial f}{\partial x} = y, \qquad \frac{\partial f}{\partial y} = x,$$

- Derivative tells the sensitivity of the whole expression on its variable value
  - x increase by h → $f(x,y)$ decrease by $-\ 3 * h$
  - y increase by h → $f(x,y)$ increase by $4 * h$

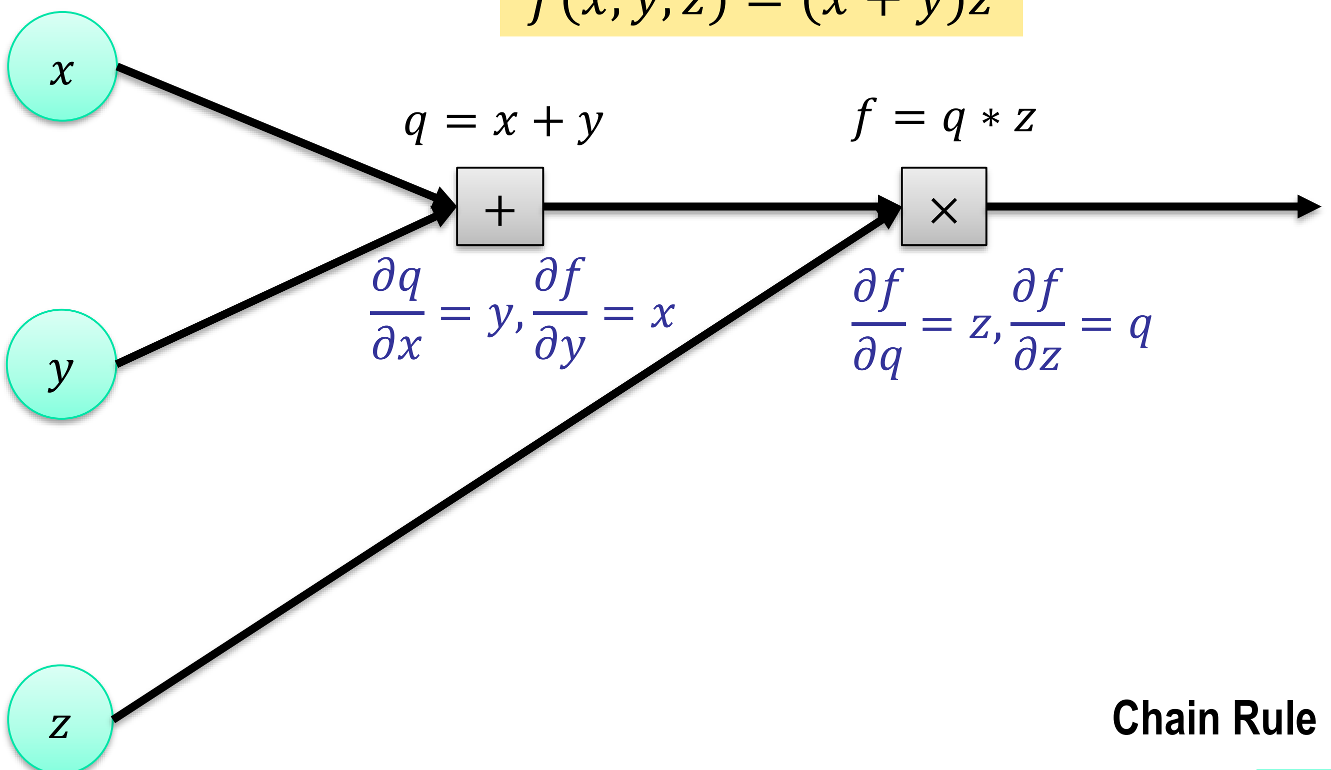
$$x = 4, y = -3, \qquad \frac{\partial f}{\partial x} = -3, \frac{\partial f}{\partial y} = 4,$$
$$f(x,y) = -12,$$

When h is small, the function is well approximated by a straight line

Chain Rule

- Calculate derivative of composed functions

$$f(x, y, z) = (x + y)z$$



Chain Rule

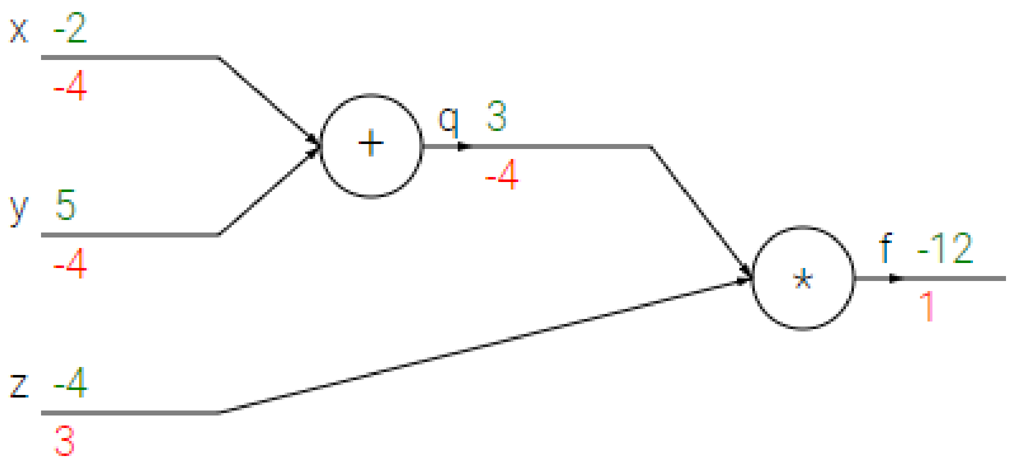
$$\frac{\partial f}{\partial z} = q = x + y$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial x} = z * y$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial y} = z * x$$

Chain Rule (2)

- Example



Logistic Regression Gradient Descent