Deep Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir

https://www.aparat.com/mehran.safayani

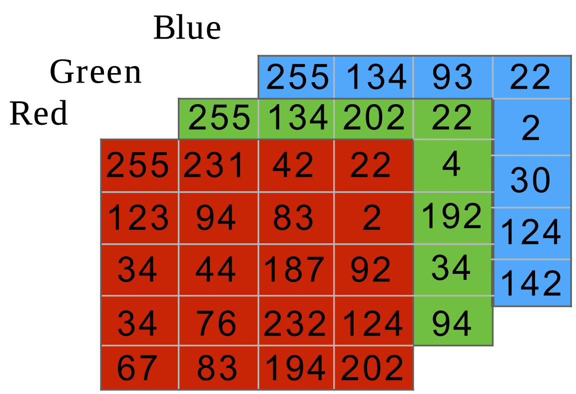
https://github.com/safayani/deep_learning_course

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

Basics of Neural Network Programming

Binary Classification

Binary classification



$$\vec{x} = \begin{bmatrix} 255 \\ 231 \\ \vdots \\ 254 \\ 253 \\ 250 \\ 220 \end{bmatrix} \qquad 64 \times 64 \times 3 = \underbrace{12288}_{n = n_x}$$

$$\vec{x} \longrightarrow \boxed{\text{model}} \longrightarrow \hat{y}$$

- Notation

$(\vec{x}, y) \qquad x \in R^{n_x}, \;\; y \in \{0,1\}$

Binary classification

- m training example: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}, \dots, (x^{(m)}, y^{(m)})\}$

- $$X = \underbrace{\begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \dots\dots\dots & x^{(m)} \\ | & | & & | \end{bmatrix}}_{n_x} \updownarrow \mathrm{m}$$

$X \in \mathbb{R}$

$X.\text{shape} = (n_x, m)$

$Y = [y^{(1)}, y^{(2)}, \dots, y^{(m)}]$

$Y \in \mathbb{R}^{1\times m}$

$Y.\text{shape} = (1, m)$

Logistic Regression

- Given x , output $\hat{y} = P(y=1|x)$ $0 \leq \hat{y} \leq 1$

$x \in \mathbb{R}^{n_x}$ parameters: $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$

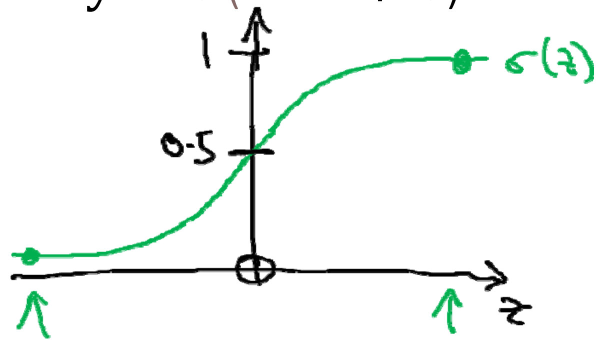
$$\hat{y} = w^T x + b$$

$$\hat{y} = \sigma(\overbrace{w^T x + b}^{z})$$

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

if z large $\sigma(z) \approx 1$

if z large negative $\sigma(z) \approx 0$



Logistic Regression

- $\hat{y} = \sigma(\underbrace{w^T x + b}_{z})$

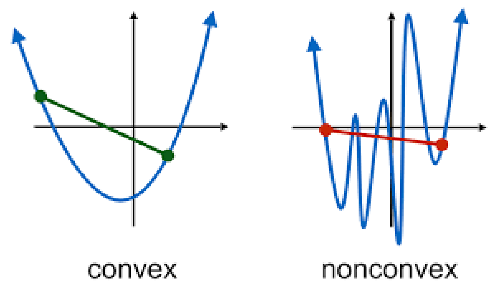
$x_0 = 1, x \in \mathbb{R}^{n_x+1}$

$$\hat{y} = w^T x$$

$$W = \begin{bmatrix} b = w_0 \\ w_1 \\ w_2 \\ w_3 \\ . \\ . \\ . \\ w_{nx} \end{bmatrix} \begin{array}{l} \} \, b \\ \\ \\ \\ \left.\begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \end{array}\right\} w \end{array}$$

Logistic Regression cost function

- Loss (error) function: $L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2 = \frac{1}{2}\left(\sigma\left(w^T x + b\right) - y\right)^2$ SE: Square Error
- Non-convex graph:

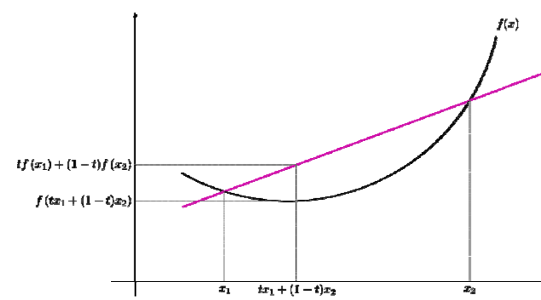


https://mlstory.org/optimization.html

- In mathematics, a real-valued function is called **convex** if the line segment between any two distinct points on the graph of the function lies above the graph between the two points.

For all $0 \leq t \leq 1$ and all $x_1, x_2 \in X$:

$$f\left(tx_1 + (1-t)x_2\right) \leq tf\left(x_1\right) + (1-t)f\left(x_2\right)$$

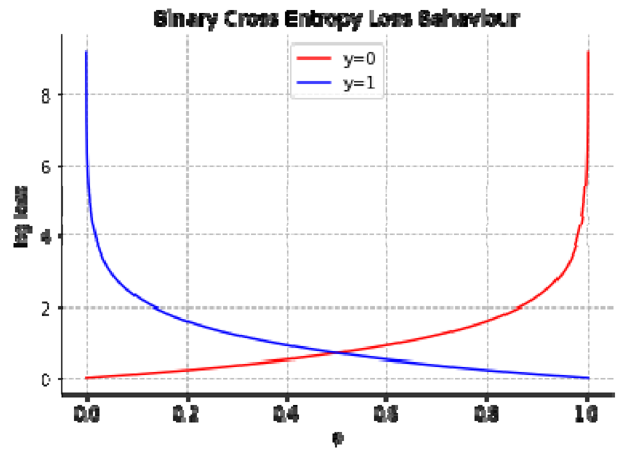


Cross Entropy

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

if $y=1$: $L(\hat{y}, y) = -\log \hat{y}$

if $y=0$: $L(\hat{y}, y) = -\log(1-\hat{y})$



https://datamonje.com/classification-loss-functions/

- Cost function: $J(w,b) = \frac{1}{m}\sum_{i=1}^{m} L\left(y^{(i)}, \hat{y}^{(i)}\right)$

$$= -\frac{1}{m}\sum_{i=1}^{m} [y^{(i)}\log \hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})]$$