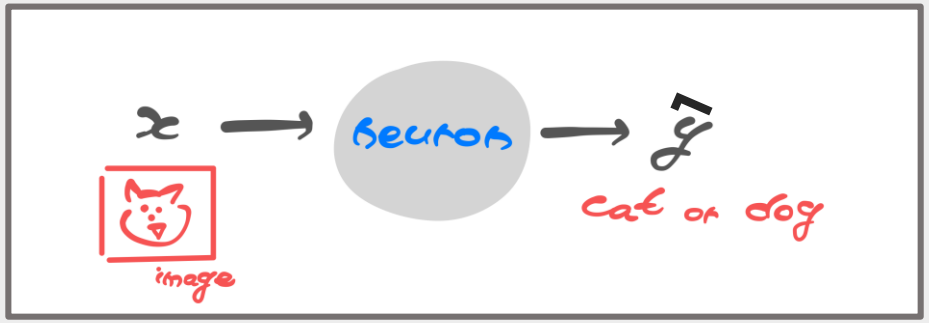
Neural Network

신경망

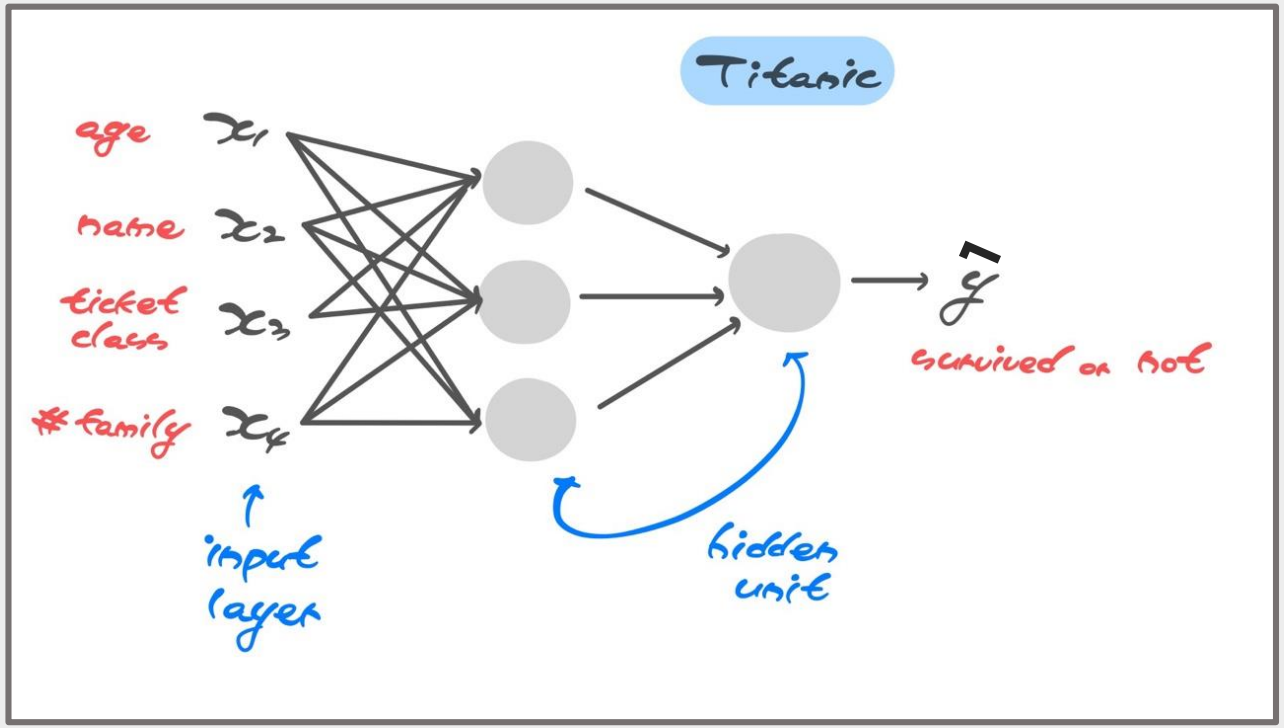
수학과 오서영

What is Neural Network?

Neural Network
Algorithm inspired by how the brain works
The role of neural network is to predict y_hat (We need to give the input x and output y)



Single neural network



Multiple neural network

Binary Classification with Logistic Regression

Logistic Regression
Learning algorithm used for classification of 2 classes.
-> output y are all either zero or one

① input $x \in \mathbb{R}^{n_x}$ where n_x is # features

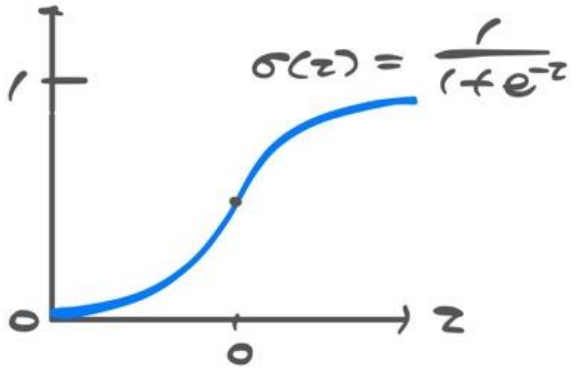
② training label $y \in \{0, 1\}$

③ parameters : $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$

④ Output $\hat{y} = \sigma(w^T x + b)$

⑤ sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$

if z is large positive, $\sigma(z) \cong 1$
if z is large negative, $\sigma(z) \cong 0$
if $z = 0$, $\sigma(z) = 0.5$



Binary Classification with Logistic Regression

Cost Function
: average of the loss function of the entire training set
For training W, b (parameters), we need to define cost function

Loss Function
Loss function computes the error between prediction y_hat and desired output y

$$L(\hat{y}^{(i)}, y^{(i)}) = -\left(y^{(i)} \log(\hat{y}^{(i)}) + (1-y^{(i)}) \log(1-\hat{y}^{(i)})\right)$$

① If $y^{(i)} = 1$, $L(\hat{y}^{(i)}, y^{(i)}) = -\log(\hat{y}^{(i)})$

② If $y^{(i)} = 0$, $L(\hat{y}^{(i)}, y^{(i)}) = -\log(1-\hat{y}^{(i)})$

Loss Function
(Cross entropy)

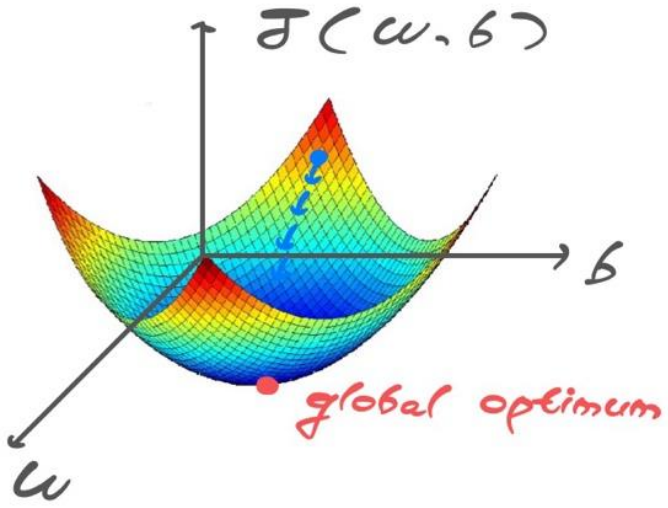
$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$$

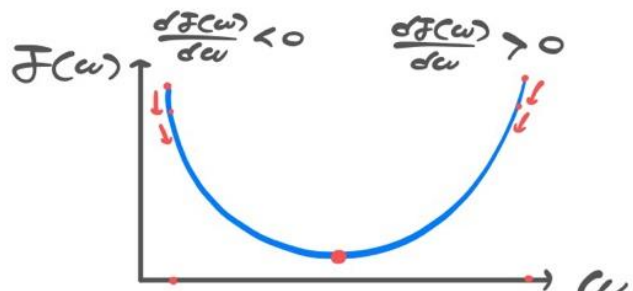
$$= -\frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} \log(\hat{y}^{(i)}) + (1-y^{(i)}) \log(1-\hat{y}^{(i)})\right)$$

Cost Function

How to minimize cost function?

Gradient Descent
Want to find W, b that minimize J(W, b)



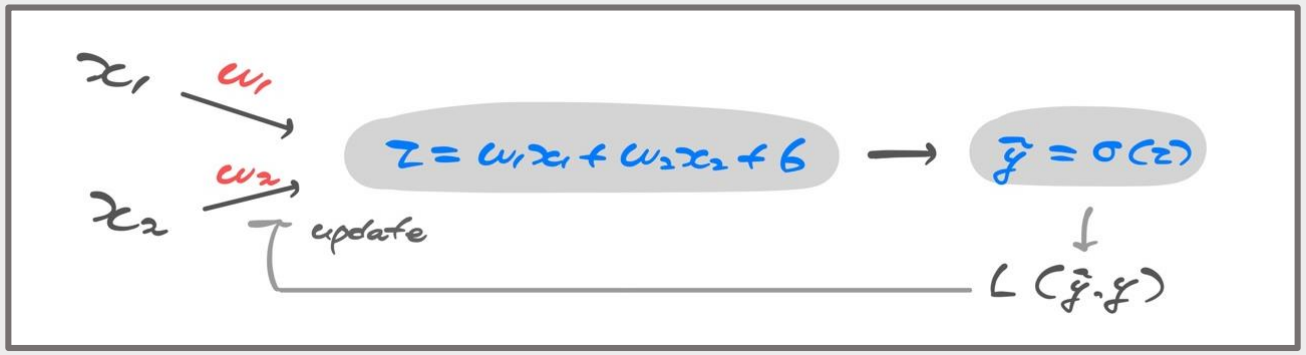


$$w := w - \alpha \frac{\partial J(w,b)}{\partial w} \quad \text{(update)}$$

$$b := b - \alpha \frac{\partial J(w,b)}{\partial b}$$

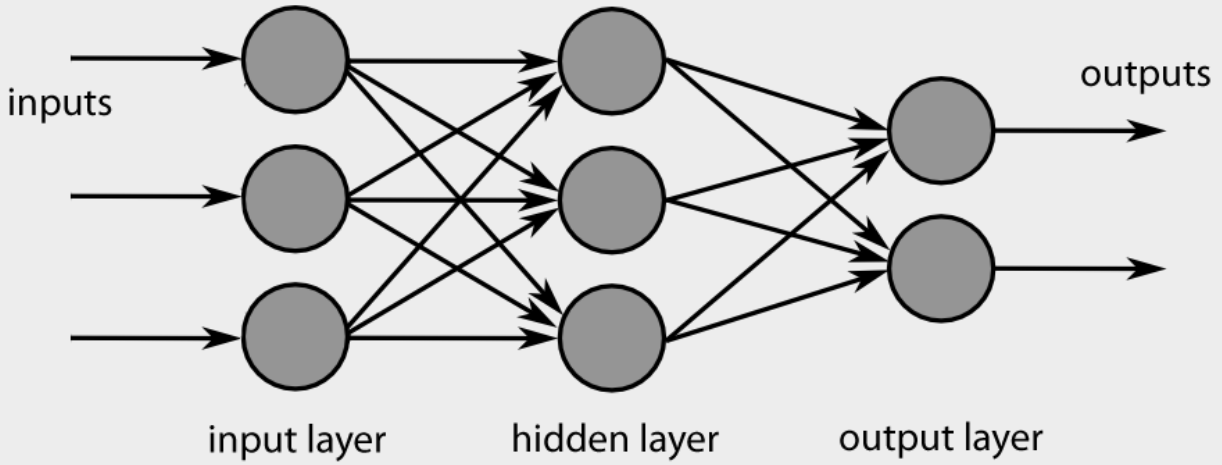
α : learning rate

Logistic Regression vs Neural Network



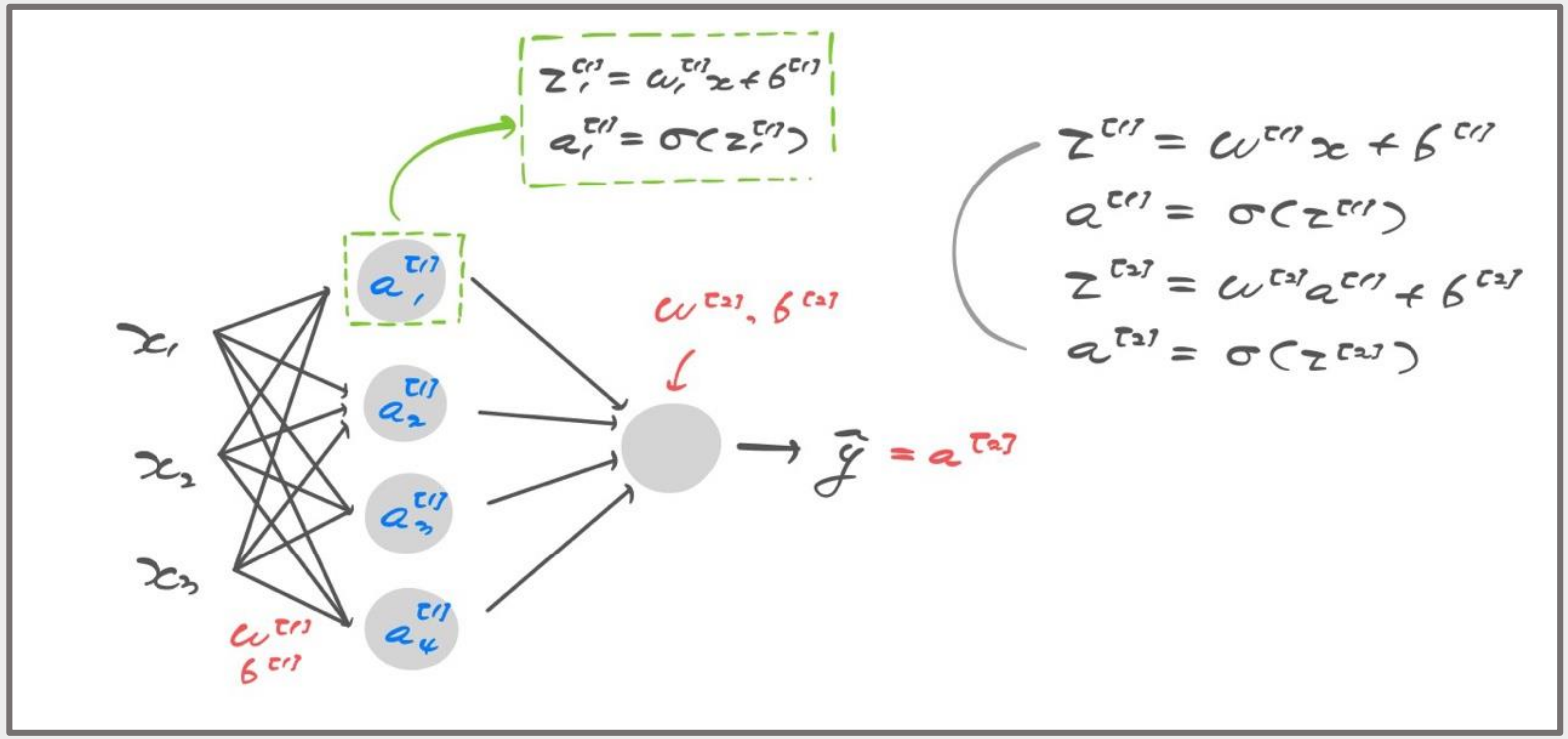
Generalization

No hidden unit, cross entropy, sigmoid function



Neural Network

Logistic Regression vs Neural Network



① sigmoid : $a = \frac{1}{1+e^{-z}}$

② tanh : $a = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

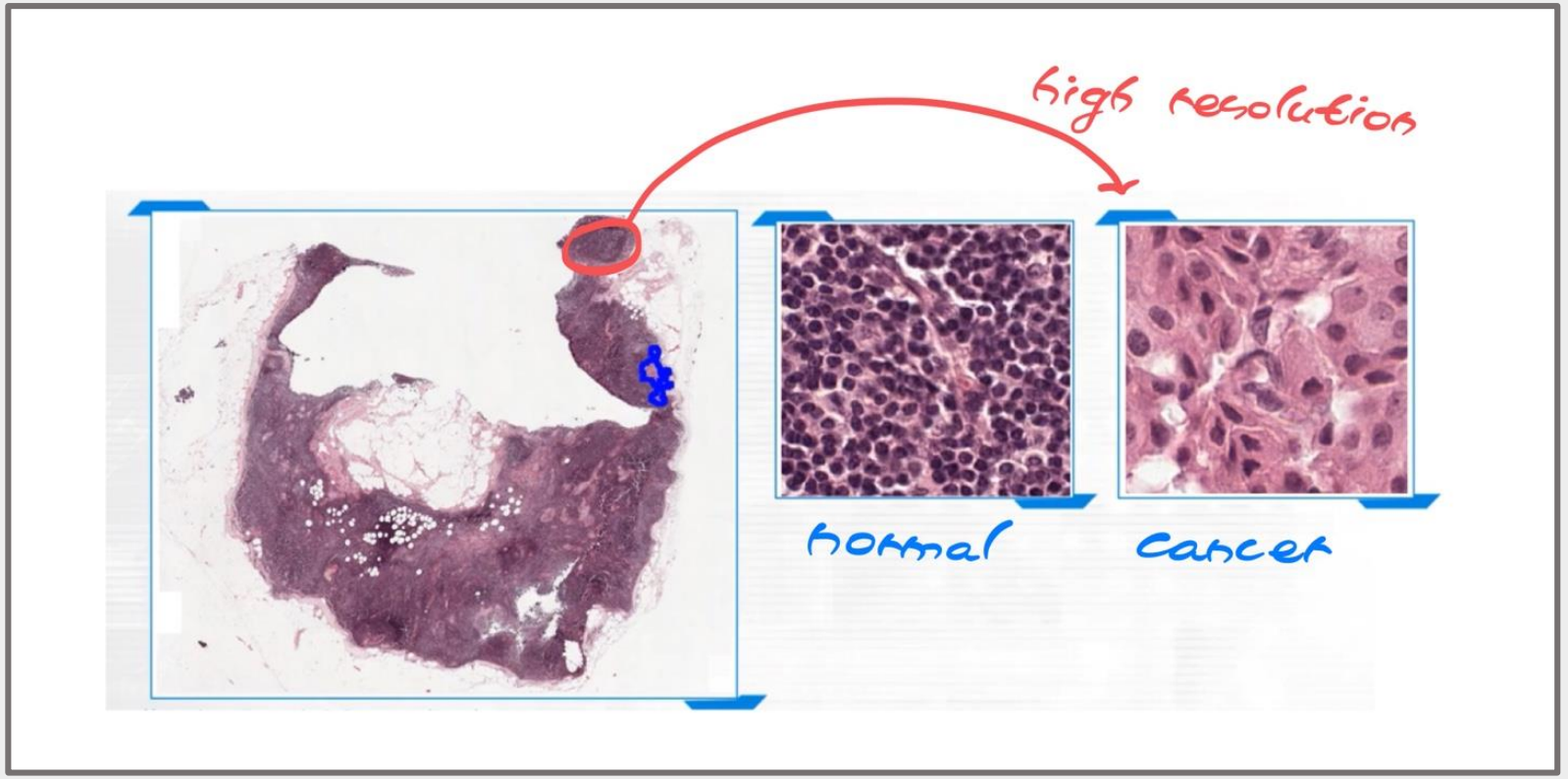
③ ReLU : $a = \max(0, z)$

→ Nonlinear

: linear hidden layer is useless

Activation function

Applications of neural networks



Pathology image classification