Deep learning

- computational power
- data available
- Algorithms

바탕이되어서 발전

Logistic Regression

$\boxed{\text{ex) Find cat}}$

i) initialize w, b

ii) Find the optimal w, b

iii) Use $\hat{y} = \sigma(wx+b)$ to predict

$\rightarrow$ loss $= -[y\log\hat{y} + (1-y)\log(1-\hat{y})]$

Gradient Descent

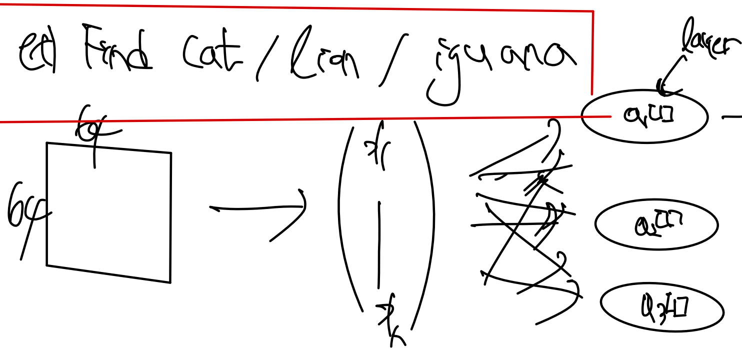
$$w = w - \alpha\frac{df}{dw}$$

$$b = b - \alpha\frac{df}{db}$$

(Eq) neuron = linear + activation

(Eq) model = architecture + parameter

$\boxed{\text{ex) Find cat / lion / iguana}}$

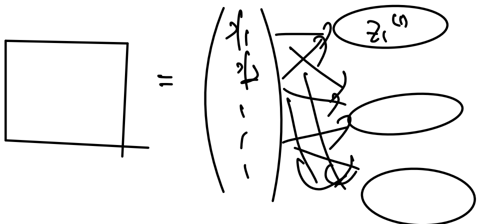


$$\rightarrow \hat{y}_1 = a_1^{[1]} = \sigma(\underbrace{w_1^{[1]}x + b_1^{[1]}}_{z^{(1)}})$$

* 정답의 data가 충분하면 뉴런이 자신이 찾을 것을 결이 무엇인지 이해한다. → Robust 함

→ But 문제점 존재함. 특징량을 분류시 영향력 ↓ 유관의 관계 X

Ex) 조건 (unique animal on an image)



$= e^{z_1^{[1]}} / \sum e^{z_k^{[1]}}$

$=$ 〃

$=$ 〃

3 class의 합이 1

* softmax를 적용함으로써 확률 분포로 표현하는 softmax multi-class regression. 이제 매개변수 training에 대해서 생각!

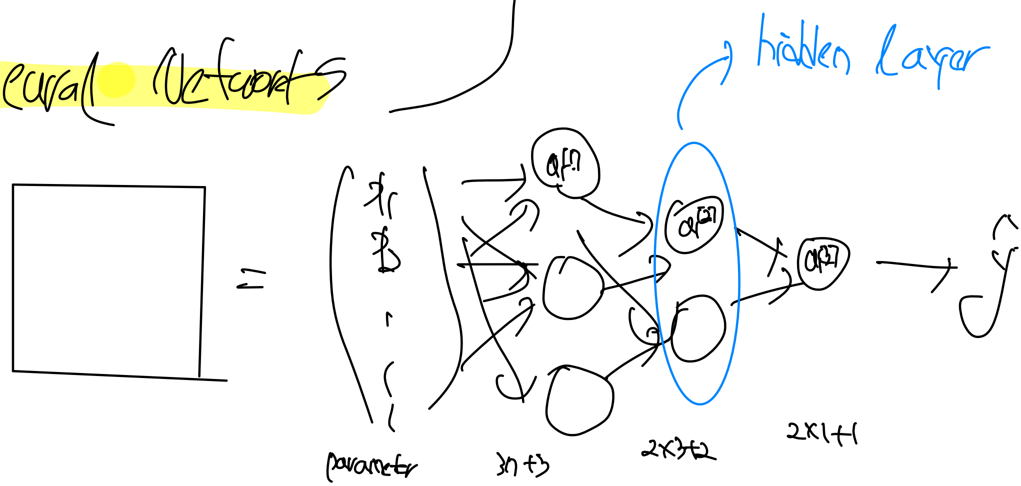
* Cross-entropy loss

$$\boxed{\mathcal{L}_{CE} = -\sum_{k=1}^{3} y_k \log \hat{y}_k}$$

end to end learning

black box model

Neural Network



Propagation equations

$$X = \begin{pmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ | & | & & | \end{pmatrix}$$ m examples ..

id of example

Solve by Broadcasting

$(3,1) \longrightarrow (3,m)$

$$b^{[1]} = \begin{pmatrix} | & | & \\ b^{[1]} & b^{[1]} & \cdots \\ | & | & \end{pmatrix}$$

$$z^{[1]} = w^{[1]} x + b^{[1]}$$

$$(3,1) = (3,n)(n,1) + (3,1) \xrightarrow{\text{m개 ex.}} (3,m) = (3,n)(n,m) + \underline{(3,1)}$$

?

* 어떤 Architecture 사용할지는 경험 & 직관에 의해 결정.

Optimizing $w^{[1]}, w^{[2]}, w^{[3]}, b^{[1]}, b^{[2]}, b^{[3]}$

define loss/cost function

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

batch 경사하강법

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

Backward Propagation

하고싶은 것은, $\forall\, l = 1, 2 \ldots 3$

$$\begin{cases} w^{[l]} = w^{[l]} - \alpha \dfrac{\partial J}{\partial w^{[l]}} \\ b^{[l]} = b^{[l]} - \alpha \dfrac{\partial J}{\partial b^{[l]}} \end{cases}$$

$$\frac{dy}{dw^{[3]}} \xrightarrow[\text{rule}]{\text{chain}} \frac{dy}{dw^{[2]}}$$

$$\frac{dy}{dw^{[3]}} = \frac{dy}{da^{[3]}} \cdot \frac{da^{[3]}}{dz^{[3]}} \cdot \frac{dz^{[3]}}{dw^{[3]}}$$

$$\frac{dy}{dw^{[2]}} = \frac{dy}{dz^{[3]}} \cdot \frac{dz^{[3]}}{da^{[2]}} \cdot \frac{da^{[2]}}{dz^{[2]}} \cdot \frac{dz^{[2]}}{dw^{[2]}}$$