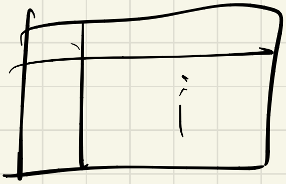
Structured Data vs Unstructured Data



- ✓ Audio
- ✓ Image
- ✓ Text

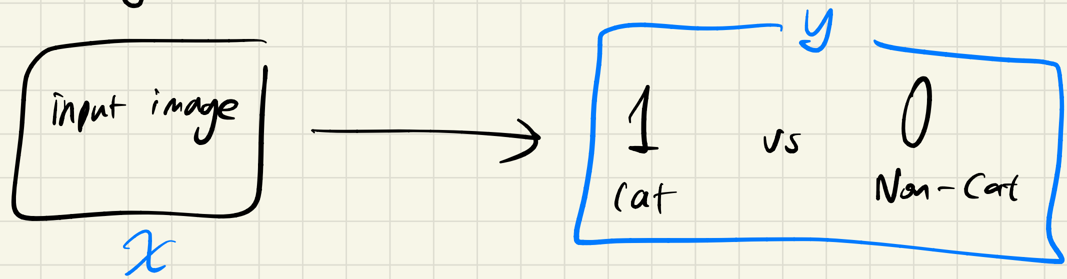
---

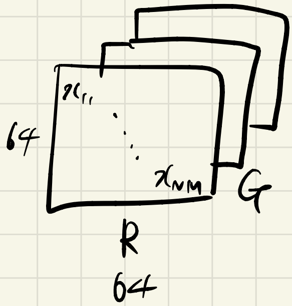
1) Data
2) Computation
3) Algorithms

Idea

Experiment ← Code

[Binary Classification]





$$\rightarrow \quad x = \begin{bmatrix} x_{11} \\ \vdots \\ x_{NM} \end{bmatrix}$$

$$64 \times 64 \times 3 = \underbrace{12{,}288}_{\text{dimension}}$$

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

$$m = m\_train$$

$$X = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ | & | & & | \end{bmatrix} \updownarrow n_x \qquad X \in \mathbb{R}^{n_x \times m}$$

$\longleftarrow m \longrightarrow$

X.shape

$$Y = [y^{(1)}, y^{(2)}, \ldots, y^{(m)}] \qquad Y \in \mathbb{R}^{1 \times m}$$

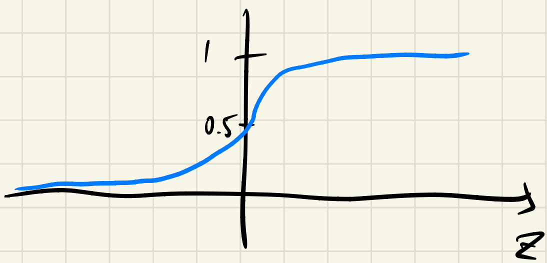
Y.shape

[Logistic Regression]

$$\hat{y} = P(y=1 \mid x)$$

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

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



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

다른 표기법

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

$$\hat{y} = \sigma(\theta^T x)$$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n_x} \end{bmatrix}$$

$\theta_1 \rightarrow b$

$\theta_2, \dots, \theta_{n_x} \rightarrow w$

[Loss function]

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

$$\underbrace{L(\hat{y}, y)}_{\text{minimize}} = -(y\log\hat{y} + (1-y)\log(1-\hat{y}))$$

If $y=1$ $= -(1\cdot\log\hat{y} + 0\cdot\log(1-\hat{y}))$

$= -\log\hat{y}$ ← $\log\hat{y}$ large, $\hat{y}$ large

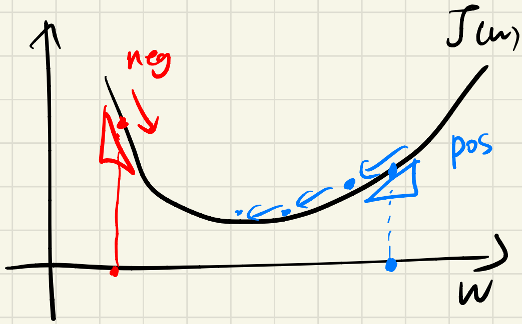
If $y=0$ $= -(0 + 1\cdot\log(1-\hat{y}))$

$= -\log(1-\hat{y})$ ← $\log(1-\hat{y})$ large, $\hat{y}$ small

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

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

[Gradient Descent]



learning rate

dw

$$w := w - \alpha \frac{dJ(w)}{dw}$$

$$= w - \alpha \cdot dw$$

∂

$$J(w, b) =$$

$$w := w - \alpha \frac{dJ(w,b)}{dw}$$

$$b := b - \alpha \frac{dJ(w,b)}{dw}$$

[Computation Graph]

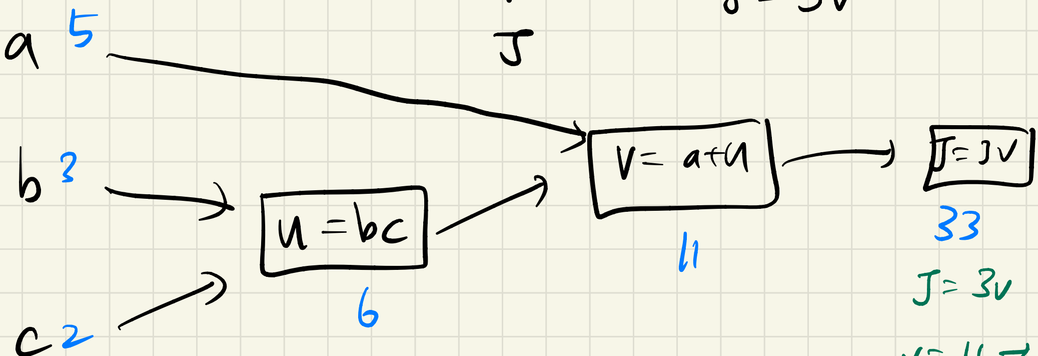
$$J(a,b,c) = 3(a + \underbrace{\underbrace{bc}_{u}}_{v})$$

(The brace labelled J spans the whole expression $3(a+bc)$.)

$u = bc$

$v = a + bc = a + u$

$J = 3v$



$J = 3v$

$v = 11 \to 11.001$

$J = 33 \to 33.003$

$$\frac{dJ}{dv} = 3$$

$a \to v \to J$

$$\frac{dJ}{da} = 3 = \frac{dJ}{\cancel{dv}} \cdot \frac{\cancel{dv}}{da} = \frac{dJ}{da}$$

$a = 5.001$

$v = 11.001$

$J = 33.003$

$$\frac{dJ}{du} = 3 = \frac{dJ}{\cancel{dv}} \cdot \frac{\cancel{dv}}{du} = \frac{dJ}{du}$$

$u = 6 \to 6.001$

$v = 11.001$

$J = 33.003$

$$\frac{dJ}{db} = \underbrace{\frac{dJ}{du}}_{3} \cdot \underbrace{\frac{du}{db}}_{2} = 6$$

$b = 3 \to 3.001$

$u = 6.002$

$v = 11.002$

$J = 33.006$

[LR _ GD]

$$z = w^T x + b$$

$$\hat{y} = a = \sigma(z)$$

$$L(a, y) = -(y \log(a) + (1-y) \log(1-a))$$

x_1

w_1

x_2

w_2

b

$$\boxed{z = w_1 x_1 + w_2 x_2 + b}$$

$$\frac{dL}{da} \cdot \frac{da}{dz}$$

$$\text{"}dz\text{"} = \frac{dL}{dz} = a - y$$

$$\boxed{\hat{y} = a = \sigma(z)}$$

$$\frac{dL}{da} = -\left(\frac{y}{a}\right) + \frac{1-y}{1-a}$$

"da"

$$\boxed{L(a, y)}$$

$$\frac{dL}{dw_1} = \text{"}dw_1\text{"} = x_1 \cdot dz$$

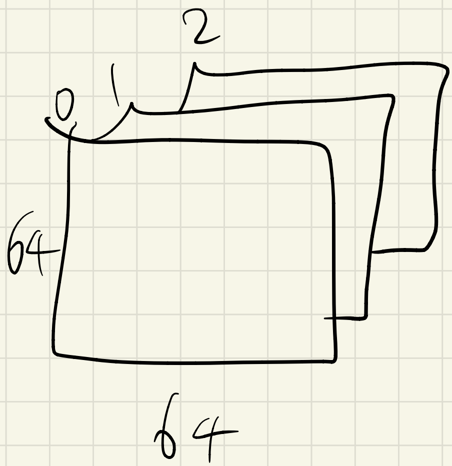
$$\text{"}dw_2\text{"} = x_2 \cdot dz$$

$$\text{"}db\text{"} = dz$$

update
$$\begin{cases} w_1 = w_1 - \alpha\, dw_1 \\ w_2 = w_2 - \alpha\, dw_2 \\ b = b - \alpha\, db \end{cases}$$

$$J(w,b) = \frac{1}{m}\sum \mathcal{L}(a,y)$$

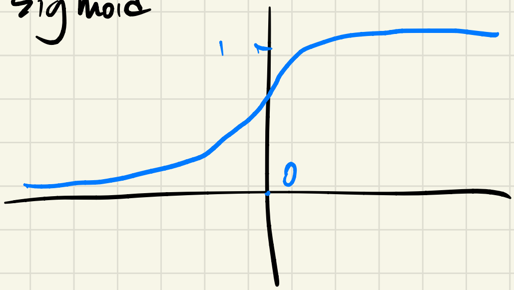
$$\frac{d}{dw} J(w,b) = \frac{1}{m}\sum \frac{d}{dw}\mathcal{L}(a,y)$$



× 209장

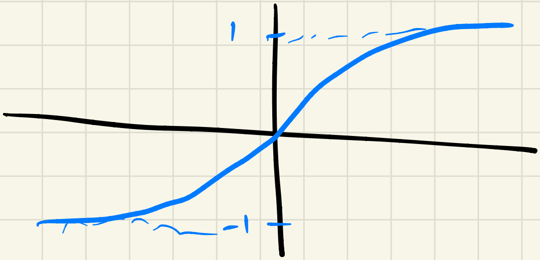
[Activation Func]

- Sigmoid bi-class



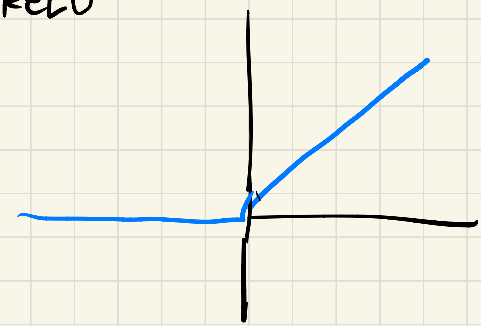
$$a = \frac{1}{1+e^{-z}}$$

- tanh



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

- ReLU mult-class



$$a = \max(0, z)$$

- Leaky ReLU

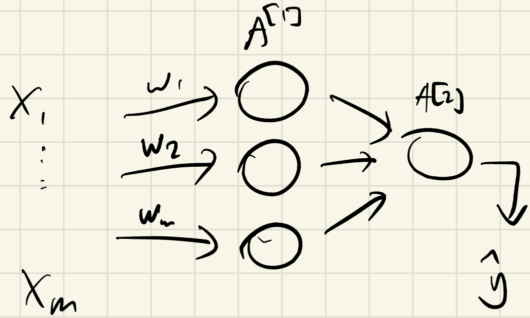
Forward Propagation

$$Z^{[1]} = W^{[1]} X + b^{[1]}$$

$$A^{[1]} = g^{[1]}(Z^{[1]})$$

$$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$$

$$A^{[2]} = g^{[2]}(Z^{[2]})$$



Back Propagation

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

$$dZ^{[2]} = A^{[2]} - Y$$

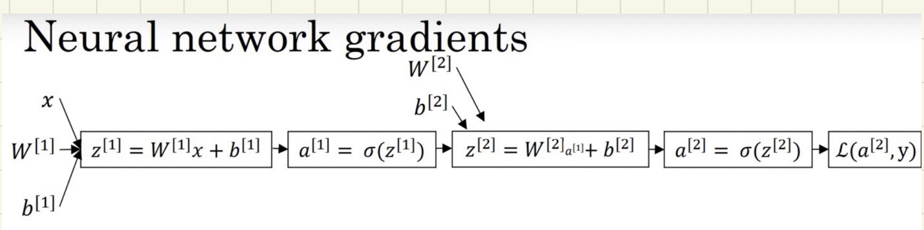
$$dW^{[2]} = \frac{1}{m} \cdot dZ^{[2]} \cdot A^{[1]T}$$

$$db^{[2]} = \frac{1}{m} \cdot \text{np.sum}(dZ^{[2]}, \text{axis}=1, \text{keepdims}=\text{True})$$

$$dZ^{[1]} = W^{[2]T} \cdot dZ^{[2]} * g^{[1]\prime}(Z^{[1]})$$

$$dW^{[1]} = \frac{1}{m} \cdot dZ^{[1]} \cdot X^T$$

$$db^{[1]} = \frac{1}{m} \text{np.sum}(dZ^{[1]}, \text{axis}=1, \text{keepdims}=\text{True})$$



Logistic regression

$$x,\ w,\ b \to \boxed{z = w^T x + b} \to \boxed{a = \sigma(z)} \to \boxed{\mathcal{L}(a,y)}$$

$dz \cdot x$ = "dw"

dz = "db"

"dz" $= \dfrac{dJ}{dz}$

$= a - y$

"da"

$= \dfrac{dJ}{da}$

$= -y \cdot \log(a) - (1-y) \cdot \log(1-a)$

$= -\dfrac{y}{a} + \dfrac{1-y}{1-a}$

$$dz = da \cdot g'(z)$$

$$g(z) = \sigma(z)$$

$$\underbrace{\frac{dJ}{dz}}_{\text{"dz"}} = \underbrace{\frac{dJ}{da}}_{\text{"da"}} \cdot \underbrace{\frac{da}{dz}}_{g'(z) =}$$

$$(\log_a x)' = \frac{1}{x \ln a}$$

$$(1-y \cdot \log(1-a))'$$

$$= \frac{1-y}{1-a \cdot \cancel{(\ln e)}} = 1$$

$x,\ W^{[1]},\ b^{[1]} \rightarrow \boxed{z^{[1]} = W^{[1]}x + b^{[1]}} \rightarrow \boxed{a^{[1]} = \sigma(z^{[1]})} \rightarrow \boxed{z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}} \rightarrow \boxed{a^{[2]} = \sigma(z^{[2]})} \rightarrow \boxed{\mathcal{L}(a^{[2]}, y)}$

$W^{[2]},\ b^{[2]} \rightarrow z^{[2]}$

$dz^{[2]} = \text{"}db^{[2]}\text{"}$

$\text{"}dW^{[2]}\text{"} = dz^{[2]} \cdot a^{[1]T}$

"$dz^{[1]}$"

"$da^{[1]}$"

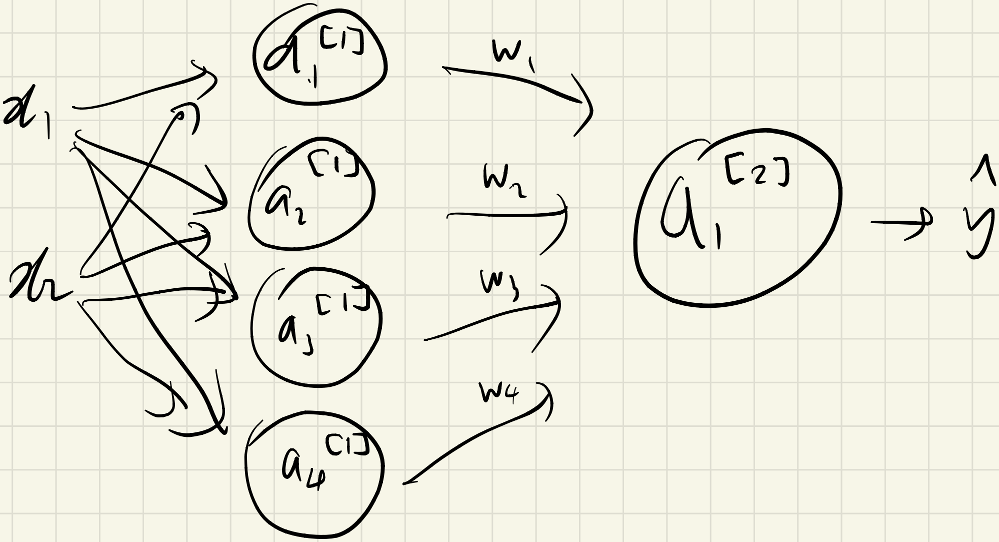
$$\text{"}dz^{[2]}\text{"} = \frac{d\mathcal{L}}{dz^{[2]}} = \frac{d\mathcal{L}}{da^{[2]}} \frac{da^{[2]}}{dz^{[2]}} = a^{[2]} - y$$

$$\text{"}da^{[2]}\text{"} = \frac{d\mathcal{L}}{da^{[2]}} = -\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}$$

$$= \frac{d\mathcal{L}}{dz^{[1]}} = \underbrace{\frac{\cancel{da^{[1]}}}{dz^{[1]}}}_{\frac{d(\sigma(z^{[1]}))}{dz^{[1]}}} \cdot \underbrace{\frac{\cancel{dz^{[2]}}}{\cancel{da^{[1]}}}}_{\frac{d(W^{[2]} \cdot a^{[1]} + b^{[2]})}{da^{[1]}}} \cdot \underbrace{\frac{\cancel{da^{[2]}}}{\cancel{dz^{[2]}}} \cdot \frac{d\mathcal{L}}{\cancel{da^{[2]}}}}_{dz^{[2]}}$$

$$\underbrace{\frac{d(\sigma(z^{[1]}))}{dz^{[1]}}}_{\sigma'(z^{[1]}) = g'(z^{[1]})} \circledast \underbrace{\frac{d(W^{[2]} \cdot a^{[1]} + b^{[2]})}{da^{[1]}}}_{W^{[2]T}} \circledast dz^{[2]}$$

$$= dz^{[2]} \cdot W^{[2]T} * g^{[1]'}(z^{[1]})$$



$$a_1^{[1]} = W_{11} \cdot x_1 + W_{21} x_2 + b_1$$

$$a_2^{[1]} = W_{12} \cdot x_1 + W_{22} x_2 + b_2$$

$$a_3^{[1]} = W_{13} \cdot x_1 + W_{23} x_2 + b_3$$

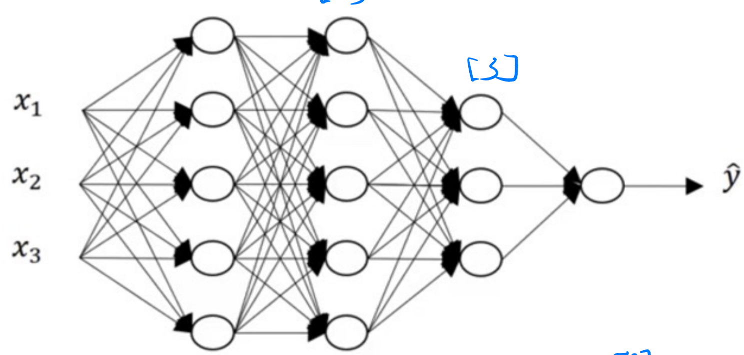
$$a_4^{[1]} = W_{14} x_1 + W_{24} x_2 + b_4$$

$$= \begin{array}{c} W_{11} \quad W_{21} \\ \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \\ W_{14} \quad W_{24} \\ (4,2) \end{array} \begin{array}{c} \begin{bmatrix} \cdot\, x_1 \\ \cdot\, x_2 \end{bmatrix} \\ (2,1) \\ n_x, m \end{array} + \begin{array}{c} \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \\ (4,1) \end{array}$$

$$a_1^{[2]} = w_1^{[2]} (W_{11} x_1 + W_{21} x_2 + b_1)$$
$$+ w_2^{[2]} (W_{12} x_1 + W_{22} x_2 + b_2)$$
$$\vdots$$
$$+ w_4^{[2]} (W_{14} x_1 + W_{24} x_2 + b_4) + b$$

$$= \underset{(1,4)}{\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \end{bmatrix}} \underset{(1,1)}{\begin{bmatrix} a_1^{[1]} \end{bmatrix}} + \begin{bmatrix} b \end{bmatrix}$$

Forward propagation in a deep network



$$x = z^{[1]} = W^{[1]} \underset{\sim a^{[0]}}{X} + b^{[1]}$$

$$a^{[1]} = g^{[1]}(z^{[1]})$$

$$z^{[2]} = W^{[2]}(a^{[1]}) + b^{[2]}$$

$$a^{[2]} = g^{[2]}(z^{[2]})$$

$\vdots$

$$z^{[4]} = W^{[4]}(a^{[3]}) + b^{[4]}$$

$$a^{[4]} = g^{[4]}(z^{[4]}) = \hat{y}$$

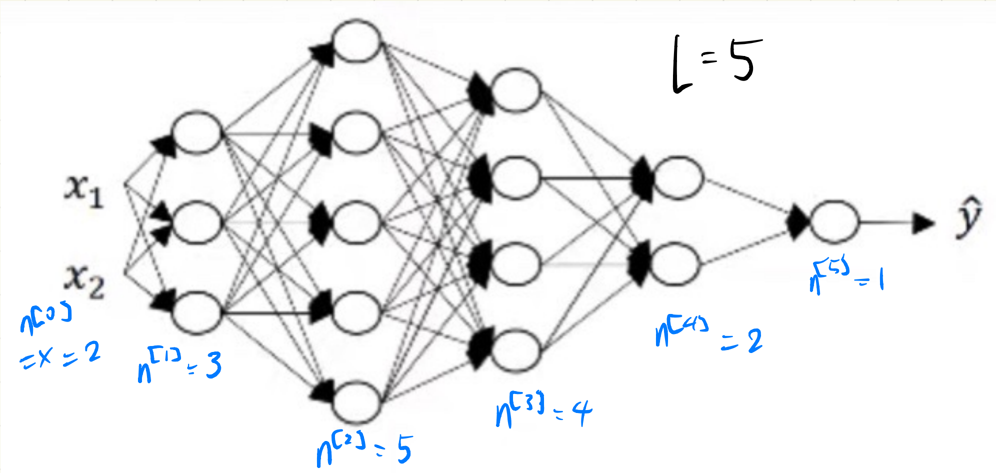
$$\therefore z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$$

$$a^{[l]} = g^{[l]}(z^{[l]})$$

Vectorize:

$$Z^{[1]} = W^{[1]} X + b^{[1]}$$

$$A^{[1]} = g^{[1]}(Z^{[1]})$$



$$W^{[l]} : (n^{[l]}, n^{[l-1]})$$

$$b^{[l]} : (n^{[l]}, 1)$$

$$dW^{[l]} : (n^{[l]}, n^{[l-1]})$$

$$db : (n^{[l]}, 1)$$

$$Z^{[1]} = W^{[1]} \cdot X + b^{[1]}$$

$(3,1)$	$(n^{[1]}, n^{[0]})$	$(2,1)$	$(3,1)$
$(n^{[1]}, 1)$		$(n^{[0]}, 1)$	$(n^{[1]}, 1)$

$$W^{[1]} : (n^{[1]}, n^{[0]})$$

$$W^{[2]} : (n^{[2]}, n^{[1]})$$

$$(5, 3)$$

$$Z^{[2]} = W^{[2]} \cdot a^{[1]} + b^{[2]}$$

$(5,1)$	$(5,3)$	$(3,1)$	$(5,1)$

$$W^{[3]} = (n^{[3]}, n^{[2]}) = (4,5)$$

$$W^{[4]} = (2,4), \quad W^{[5]} = (1,2)$$

$$Z^{[1]} = W^{[1]} \cdot X + b^{[1]}$$

$$(n^{[1]}, 1) = (n^{[1]}, n^{[0]}) \cdot (n^{[0]}, 1) + (n^{[1]}, 1)$$

vectorize

$$Z^{[1]} = W^{[1]} \cdot X + b^{[1]}$$

$$(n^{[1]}, n^{[0]}) \cdot (n^{[0]}, m) + \underbrace{(n^{[1]}, 1)}_{\text{Broadcasting} \; ; \; (n^{[1]}, m)}$$

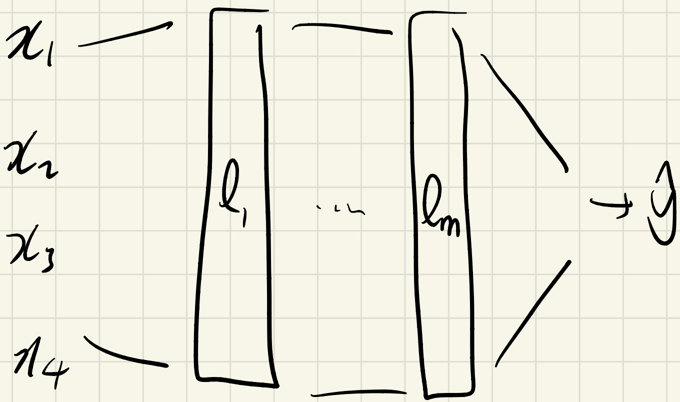
$$\begin{bmatrix} | & & | \\ Z^{[1](1)} & \cdots & Z^{[1](m)} \\ | & & | \end{bmatrix}$$

$$(n^{[1]}, \underbrace{m}_{\text{size of training set}})$$

$$Z^{[l]}, a^{[l]} : (n^{[l]}, 1)$$

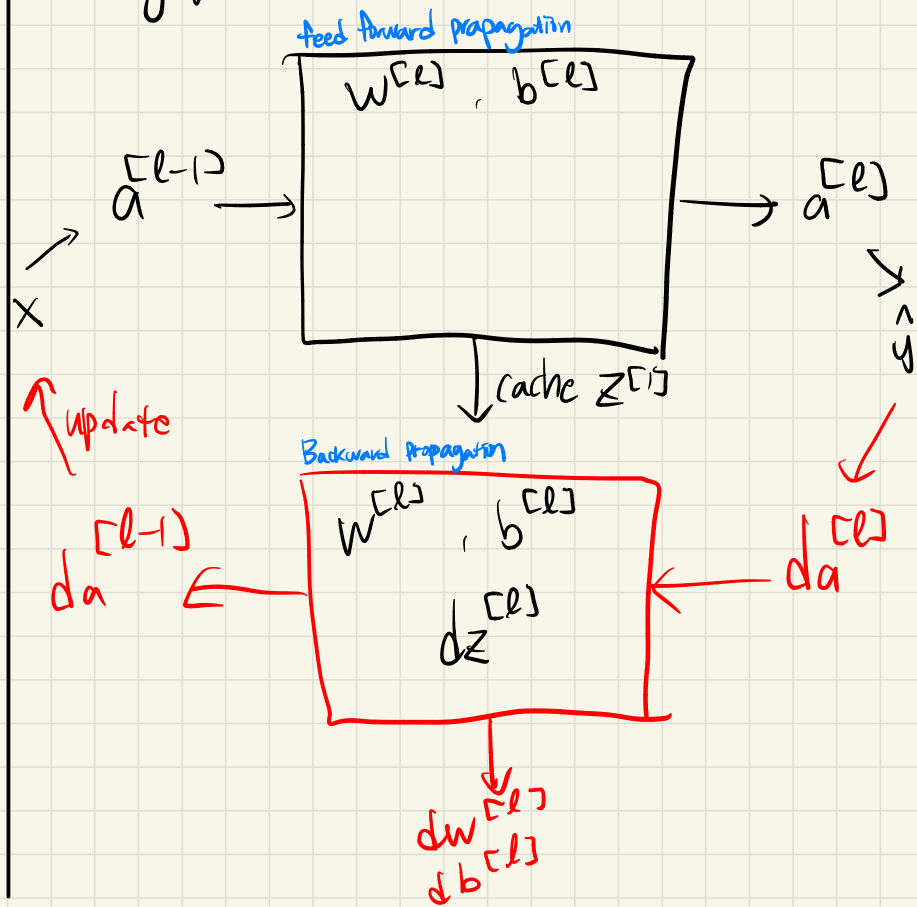
$$Z^{[l]}, A^{[l]} : (n^{[l]}, m)$$

$$l = 0 \quad A^{[0]} = X = (n^{[0]}, m)$$

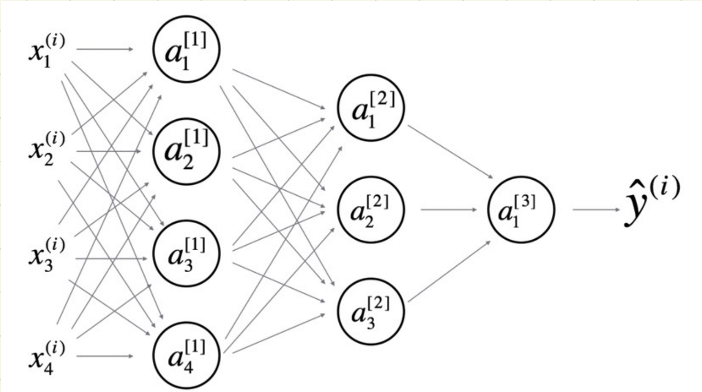


x1
x2
x3
x4
l1
...
lm
y

layer l



feed forward propagation
W[l], b[l]
a[l-1]
a[l]
x
ŷ
cache z[l]
update
Backward propagation
W[l], b[l]
dz[l]
da[l-1]
da[l]
dw[l]
db[l]



$W^{[l]} : (n^{[l]}, n^{[l-1]})$

$b^{[l]} : (n^{[l]}, 1)$

$dW^{[l]} : (n^{[l]}, n^{[l-1]})$

$db : (n^{[l]}, 1)$

$n^{[0]} = 4 \quad n^{[1]} = 4 \quad n^{[2]} = 3 \quad n^{[3]} = 1$

$\therefore W^{[1]} = (4, 4) \quad W^{[2]} = (3, 4) \quad W^{[3]} = (1, 3)$

$b^{[1]} = (4, 1) \quad b^{[2]} = (3, 1) \quad b^{[3]} = (1, 1)$