

Copyright Notice

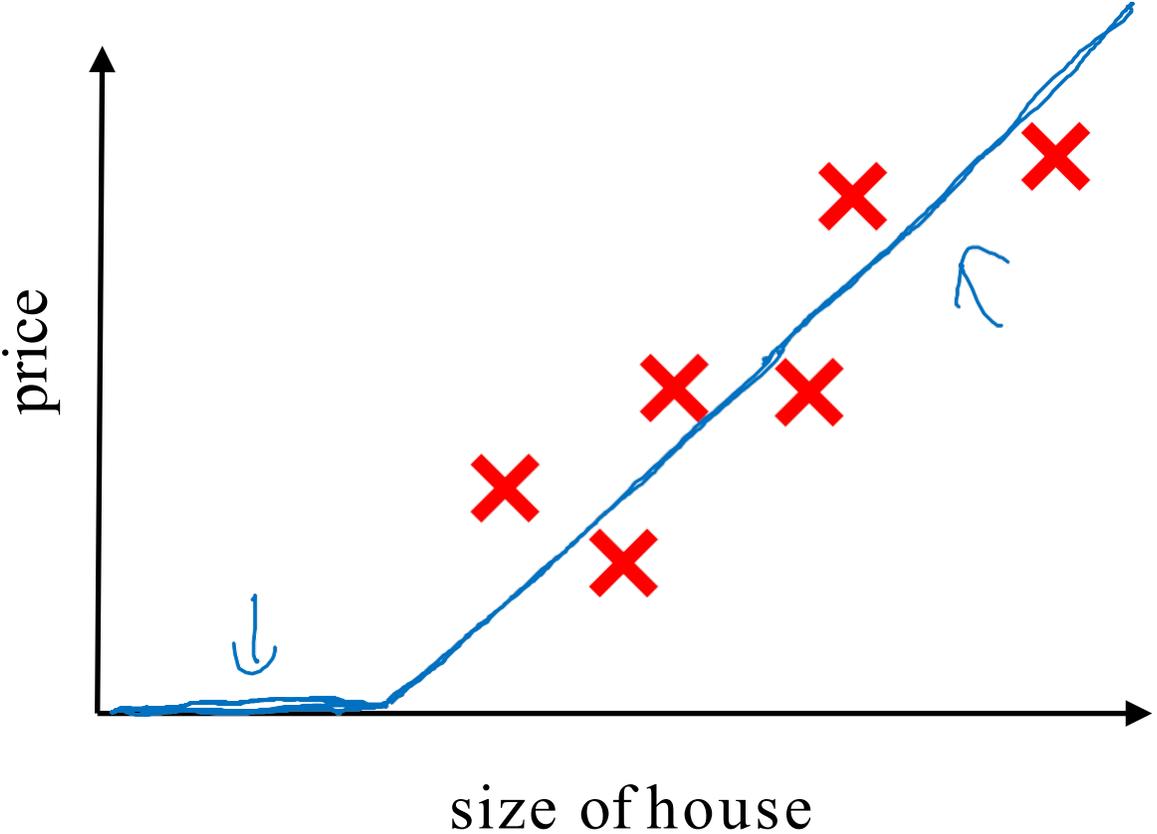
These slides are distributed under the Creative Commons License.

[DeepLearning.AI](#) makes these slides available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite [DeepLearning.AI](#) as the source of the slides.

For the rest of the details of the license, see <https://creativecommons.org/licenses/by-sa/2.0/legalcode>

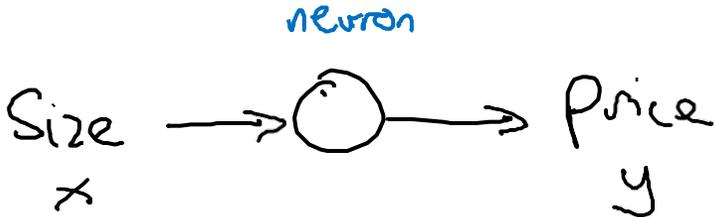
What is a
Neural Network?

Housing Price Prediction

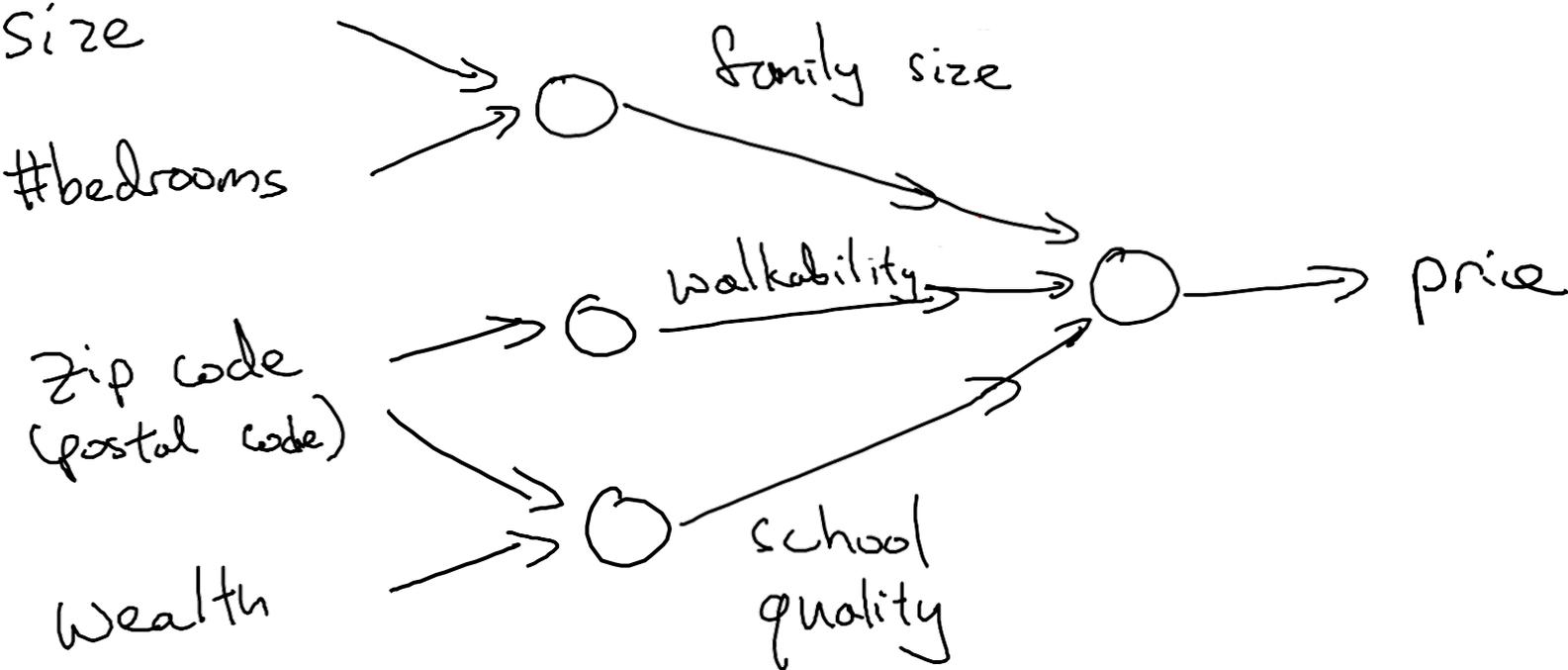


ReLU
Rectified
Linear
Unit

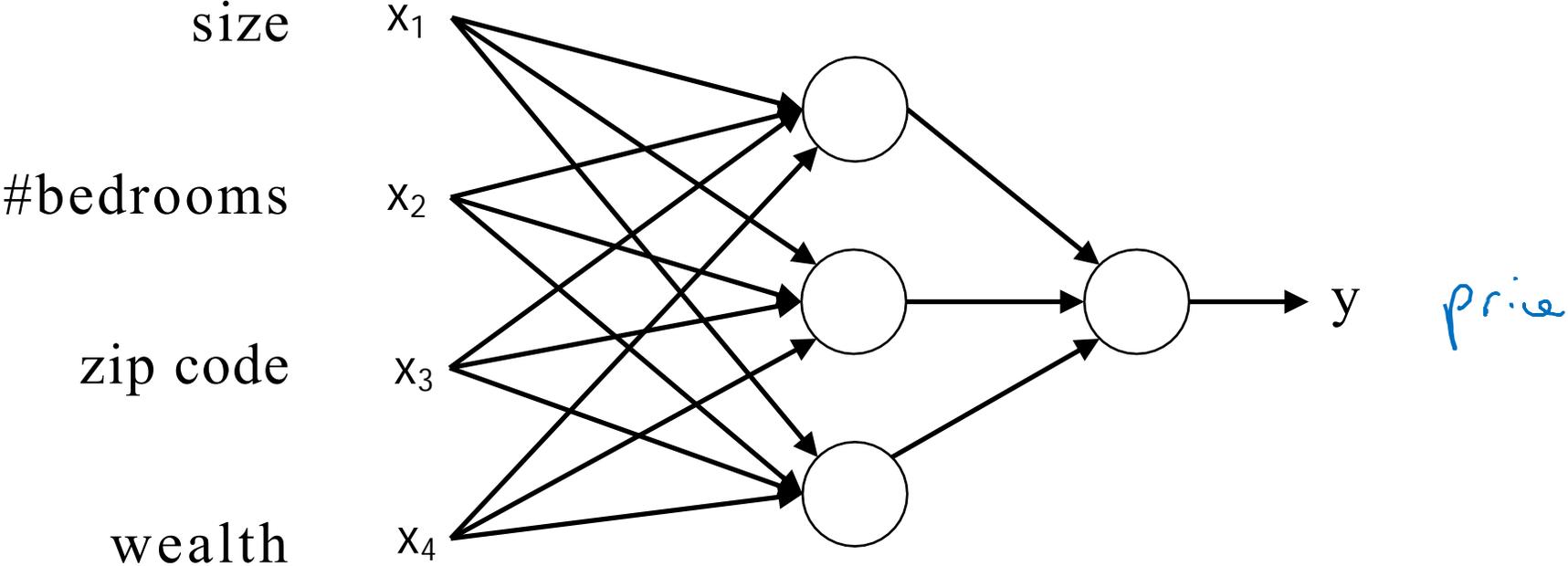
A hand-drawn graph of the ReLU function, showing a horizontal line at zero for negative inputs and a line with a positive slope for positive inputs.



Housing Price Prediction



Housing Price Prediction



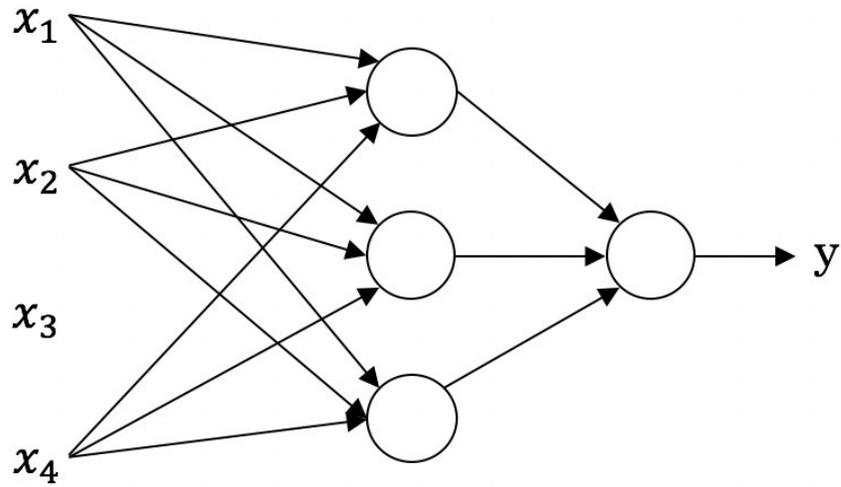
Supervised Learning

Input(x)	Output (y)	Application
Home features	Price	Real Estate
Ad, user info	Click on ad? (0/1)	Online Advertising
Image	Object (1,...,1000)	Photo tagging
Audio	Text transcript	Speech recognition
English	Chinese	Machine translation
Image, Radar info	Position of other cars	Autonomous driving

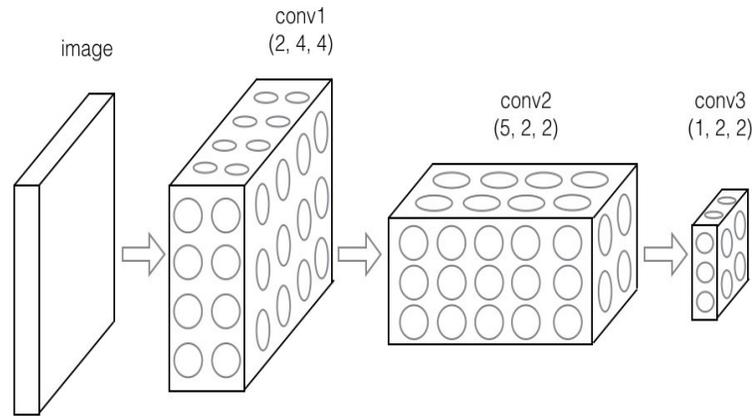
Handwritten annotations in blue ink:

- A bracket groups "Real Estate" and "Online Advertising" with the label "Standard NN".
- A bracket groups "Photo tagging" with the label "CNN".
- A bracket groups "Speech recognition" and "Machine translation" with the label "RNN".
- A bracket groups "Autonomous driving" with the label "Custom Hybrid".

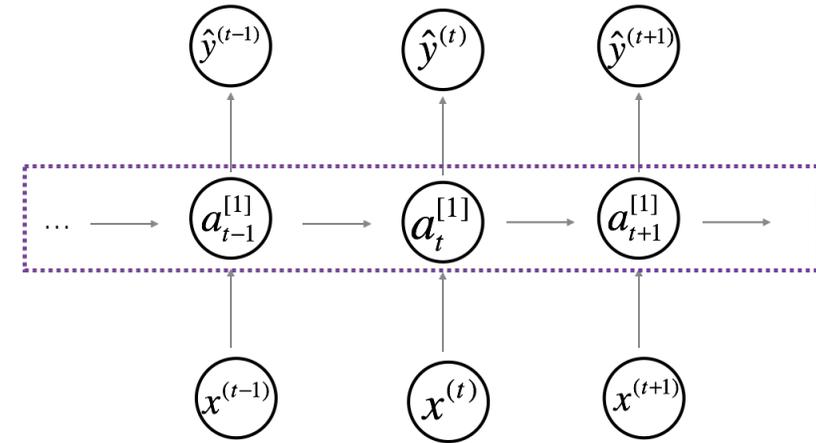
Neural Network examples



Standard NN



Convolutional NN



Recurrent NN

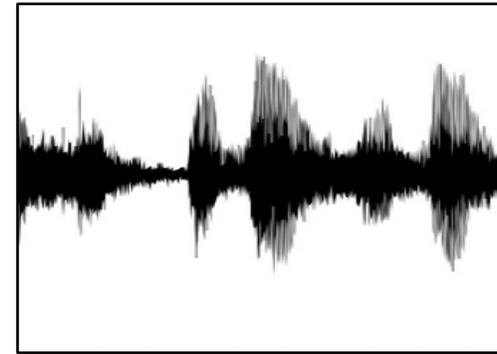
Supervised Learning

Structured Data

Size	#bedrooms	...	Price (1000\$)
2104	3		400
1600	3		330
2400	3		369
...
3000	4		540

User Age	Ad Id	...	Click
41	93242		1
80	93287		0
18	87312		1
...
27	71244		1

Unstructured Data



Audio

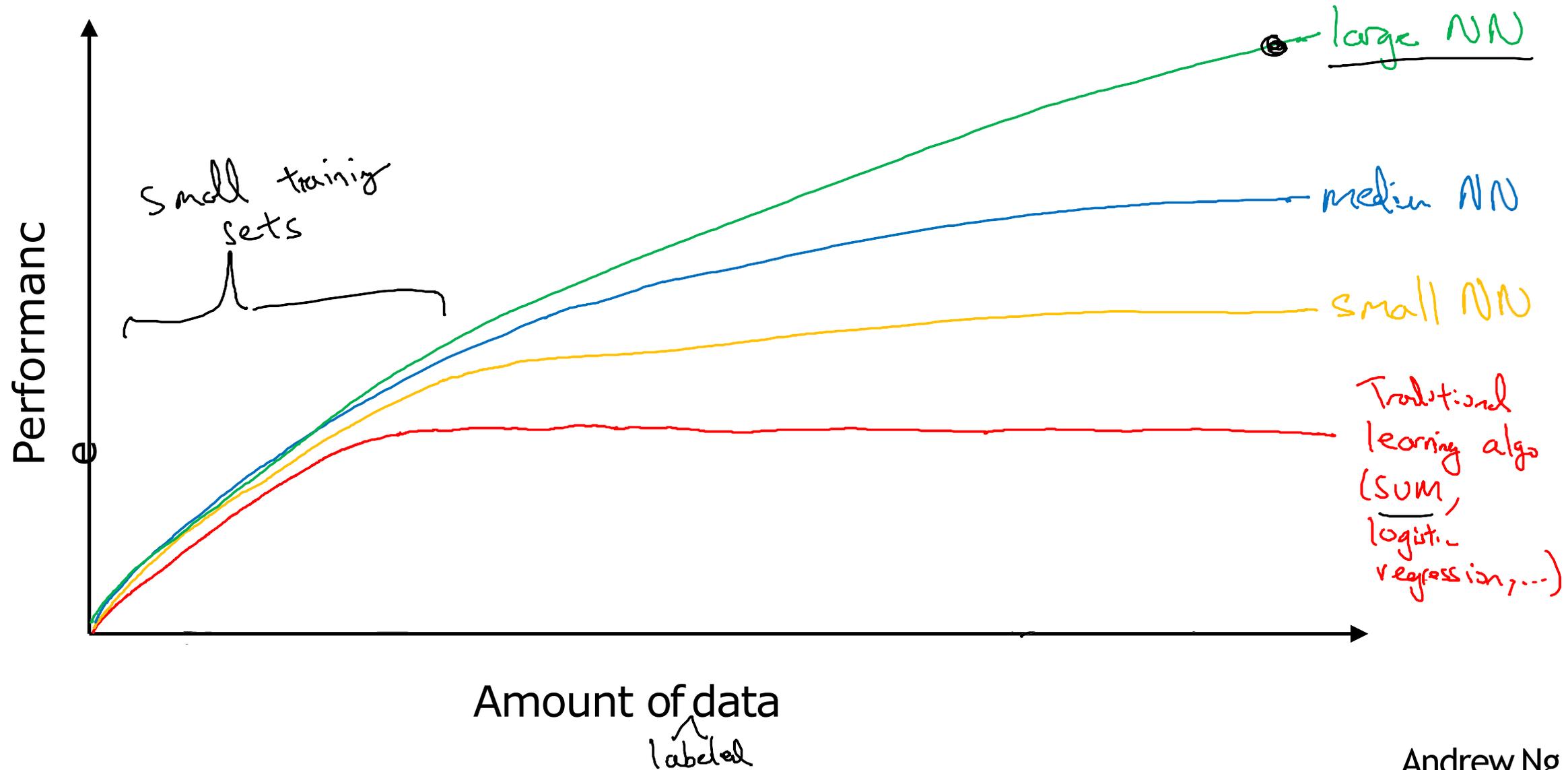


Image

Four scores and seven
years ago...

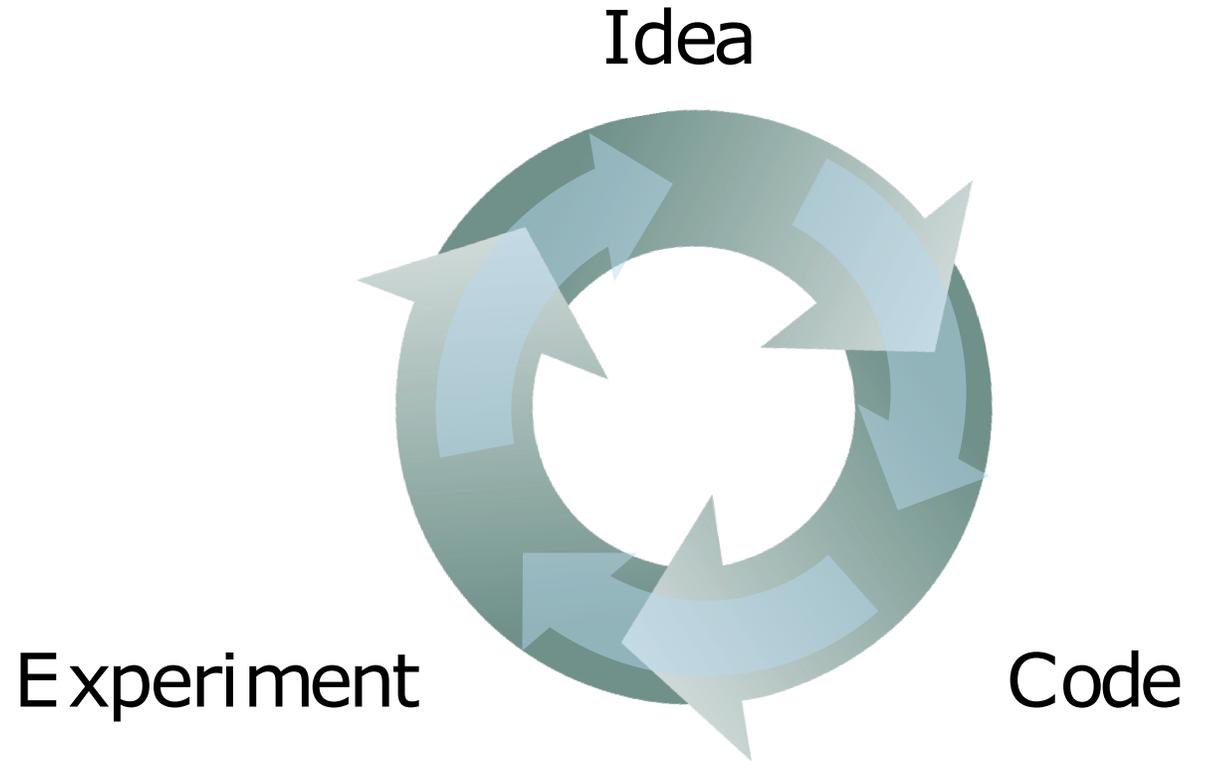
Text

Scale drives deep learning progress



Scale drives deep learning progress

- Data
- Computation
- Algorithms



Binary Classification

64



1 (cat) vs 0 (non cat)

64

Blue

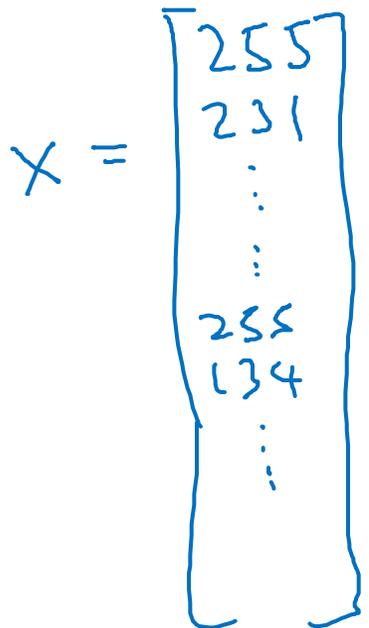
Green

Red

		255	134	93	22
	255	134	202	22	2
255	231	42	22	4	30
123	94	83	2	192	124
34	44	187	92	34	142
34	76	232	124	94	
67	83	194	202		

64

64



$$64 \times 64 \times 3 = 12288$$

$$n = n_x = 12288$$



Notation

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

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

$$M = M_{\text{train}}$$

$M_{\text{test}} = \# \text{test examples.}$

$$X = \begin{bmatrix} | & | & \dots & | \\ x^{(1)} & x^{(2)} & \dots & x^{(m)} \\ | & | & \dots & | \end{bmatrix}$$

$X \in \mathbb{R}^{n_x \times m}$ $X.\text{shape} = (n_x, m)$

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

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

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

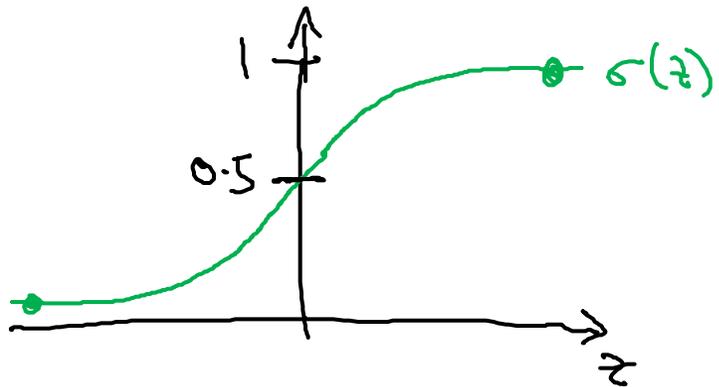
Logistic Regression

Given x , want $\hat{y} = P(y=1|x)$

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

Parameters: $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$.

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



$$x_0 = 1, \quad x \in \mathbb{R}^{n_x + 1}$$
$$\hat{y} = \sigma(\theta^T x)$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n_x} \end{bmatrix} \left. \begin{array}{l} \} b \leftarrow \\ \} w \leftarrow \end{array} \right\}$$

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

If z large $\sigma(z) \approx \frac{1}{1+0} = 1$

If z large negative number

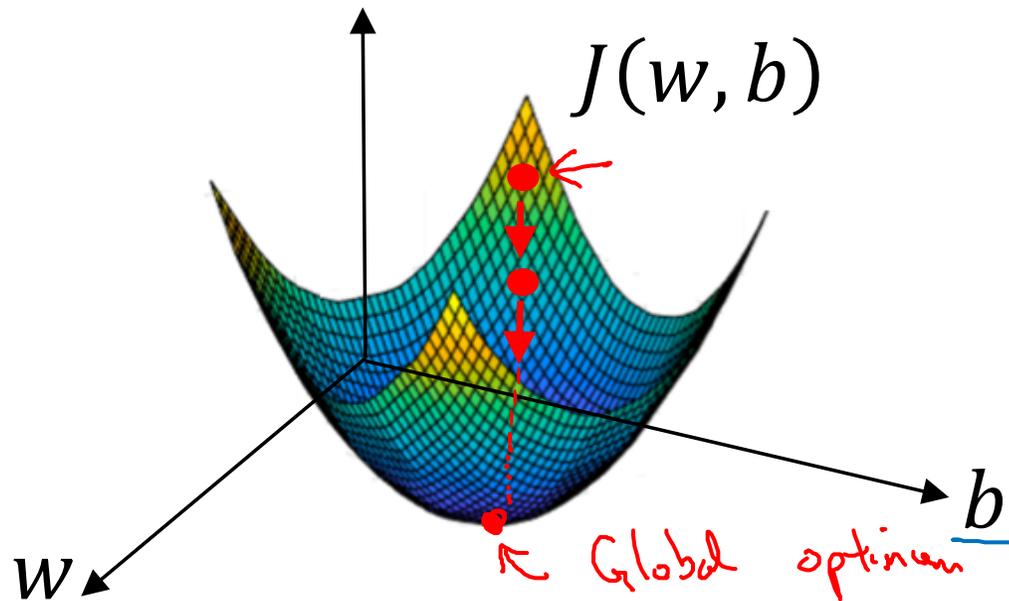
$$\sigma(z) = \frac{1}{1 + e^{-z}} \approx \frac{1}{1 + \text{Big num}} \approx 0$$

Gradient Descent

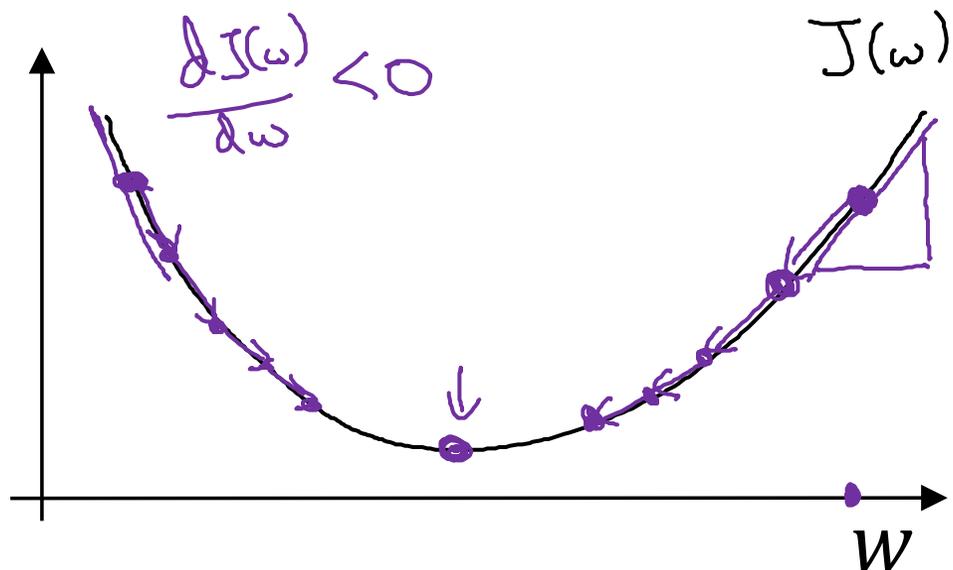
Recap: $\hat{y} = \sigma(w^T x + b)$, $\sigma(z) = \frac{1}{1+e^{-z}}$

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

Want to find w, b that minimize $J(w, b)$



Gradient Descent



Repeat {

$$w := w - \alpha \underbrace{\frac{dJ(w)}{dw}}_{\text{learning rate}}$$

}

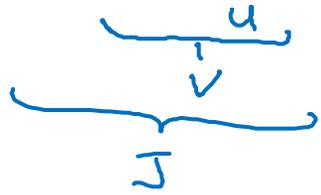
$$w := w - \alpha \underline{dw}$$

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

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

Computation Graph

$$J(a,b,c) = 3(a + \underbrace{bc}_u) = 3(5 + 3 \times 2) = 33$$



$$u = bc$$

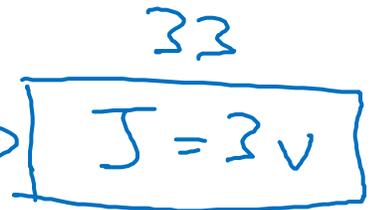
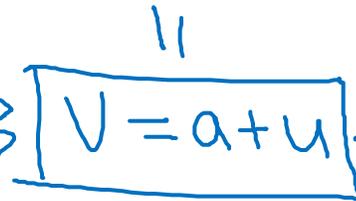
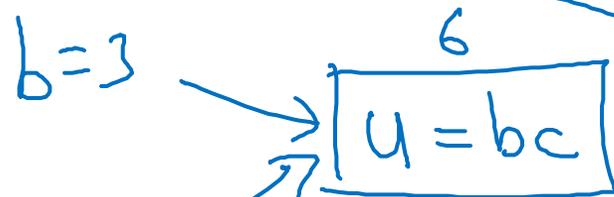
$$V = a + u$$

$$J = 3v$$

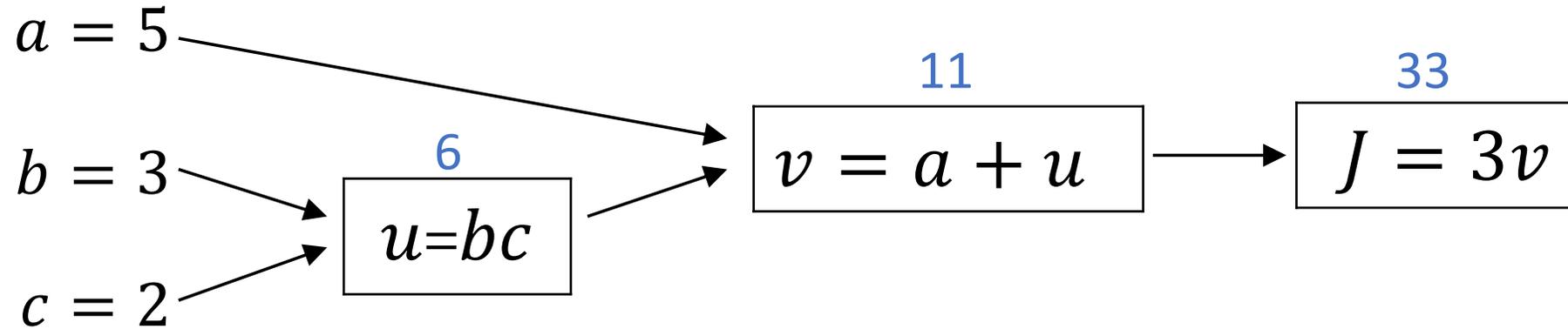
$$a = 5$$

$$b = 3$$

$$c = 2$$



Computing derivatives



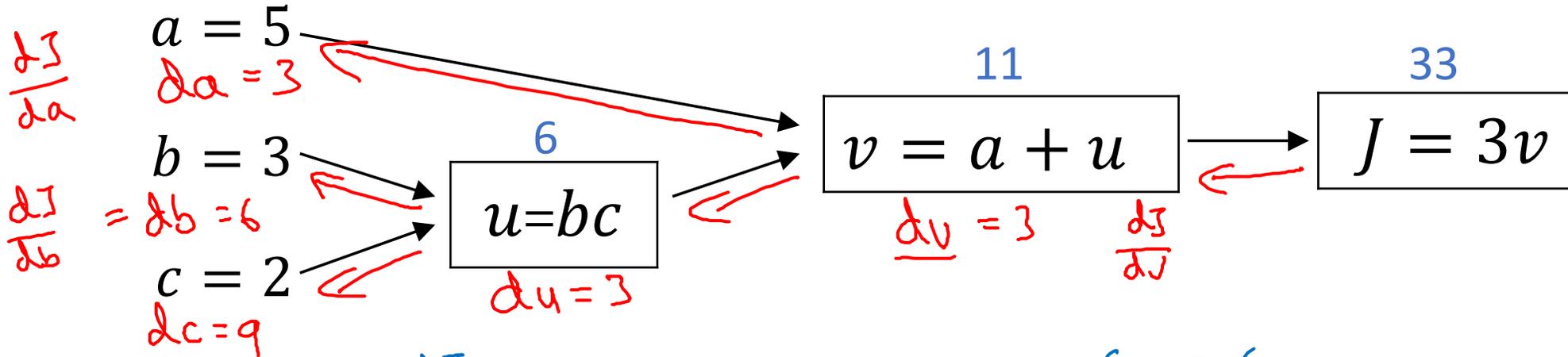
$\frac{dJ}{dv} = ? = 3$ $a \rightarrow v \rightarrow J$

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

$\frac{dv}{da} = 1$

Handwritten notes in blue and green ink showing the chain rule derivation. The first line shows $\frac{dJ}{dv} = ? = 3$ with a green arrow pointing to the second line. The second line shows $\frac{dJ}{da} = 3 = \frac{dJ}{dv} \frac{dv}{da}$ with green arrows pointing to the terms $\frac{dJ}{dv}$ and $\frac{dv}{da}$. Below the second line, the values 3×1 are written with green arrows pointing to the $\frac{dJ}{dv}$ and $\frac{dv}{da}$ terms respectively. The third line shows $\frac{dv}{da} = 1$ enclosed in a green box.

Computing derivatives



$$\frac{dJ}{du} = 3 = \underbrace{\frac{dJ}{dv}}_3 \cdot \underbrace{\frac{dv}{du}}_1$$

$$\frac{dJ}{db} = \underbrace{\frac{dJ}{du}}_3 \cdot \underbrace{\frac{du}{db}}_2 = 6$$

$$\frac{dJ}{da} = \underbrace{\frac{dJ}{du}}_3 \cdot \underbrace{\frac{du}{da}}_3 = 9$$

$$u = 6 \rightarrow 6.001$$

$$v = 11 \rightarrow 11.001$$

$$J = 33 \rightarrow 33.003$$

$$b = 3 \rightarrow 3.001$$

$$u = b \cdot c = 6 \rightarrow 6.002$$

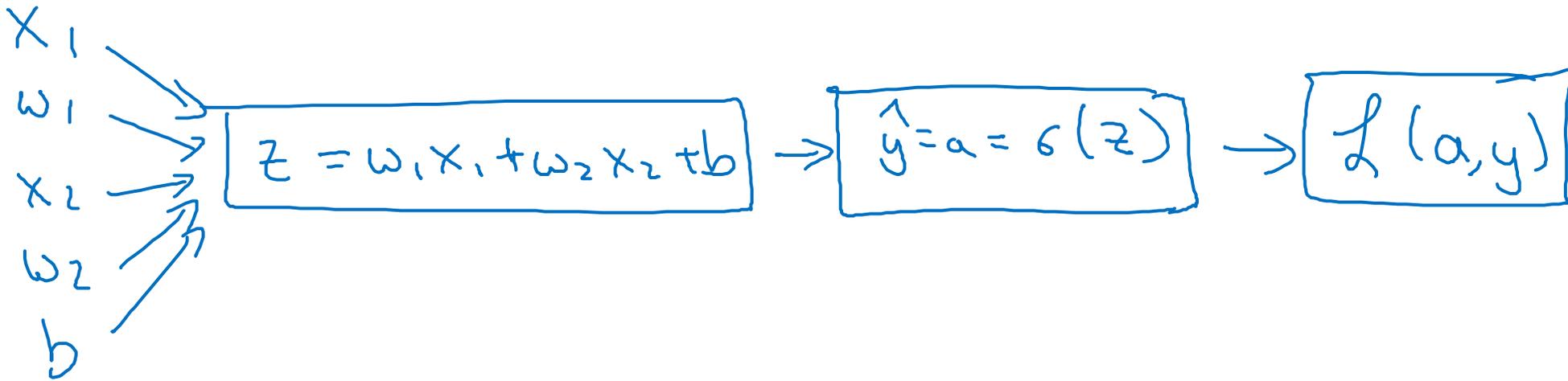
$$J = 33.006$$

Logistic regression recap

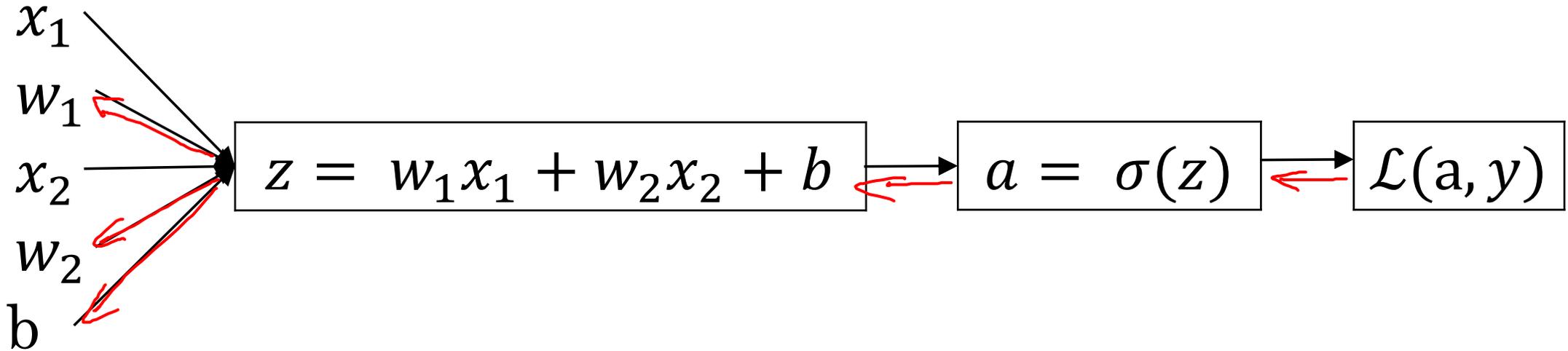
$$z = w^T x + b$$

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

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



Logistic regression derivatives



$$\sigma'(z) = \sigma(z)(1 - \sigma(z))$$

$$\frac{dL}{da} = -\frac{y}{a} + \frac{1 - y}{1 - a}$$

Logistic regression on m examples

$$\underline{J(w, b)} = \frac{1}{m} \sum_{i=1}^m \ell(a^{(i)}, y^{(i)})$$

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

$$(x^{(i)}, y^{(i)})$$

$$dw_1^{(i)}, dw_2^{(i)}, db^{(i)}$$

$$\frac{\partial}{\partial w_1} J(w, b) = \frac{1}{m} \sum_{i=1}^m \underbrace{\frac{\partial}{\partial w_1} \ell(a^{(i)}, y^{(i)})}_{dw_1^{(i)} - (x^{(i)}, y^{(i)})}$$

Implementing Logistic Regression

on m examples

$$J = 0, \quad dw_1 = 0, \quad dw_2 = 0, \quad db = 0$$

for $i = 1$ to m :

$$z^{(i)} = w^T x^{(i)} + b$$

$$a^{(i)} = \sigma(z^{(i)})$$

$$J += -[y^{(i)} \log a^{(i)} + (1 - y^{(i)}) \log(1 - a^{(i)})]$$

$$dz^{(i)} = a^{(i)} - y^{(i)}$$

$$dw_1 += x_1^{(i)} dz^{(i)}$$

$$dw_2 += x_2^{(i)} dz^{(i)}$$

$$db += dz^{(i)}$$

$$J = J/m, \quad dw_1 = dw_1/m, \quad dw_2 = dw_2/m$$

$$db = db/m$$

$$dw_1 = \frac{\partial J}{\partial w_1}$$

$$w_1 := w_1 - \alpha dw_1$$

$$w_2 := w_2 - \alpha dw_2$$

$$b := b - \alpha db$$