A Single Neuron

✓ · Data

✓ · Regression vs. Classification

· Neuron

$$\theta,\ W+b$$

$$X \longrightarrow \boxed{\text{Model}} \longrightarrow \hat{y} \quad \text{(guess)}$$

Data

$$D = \{X, Y\}$$

input

label (known)

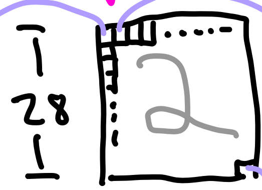
Supervised Learning

Example Problem: Image of Digit Classification

matrix

X: all of our traing images

Y: the label for each image



28

1—28—1

(784 pixels [0-255])

"2"

· Capital → matrix

· Lower case → column vector

$$x^{(i)} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_{784}^{(i)} \end{bmatrix} \qquad X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_{784}^{(1)} \\ x_1^{(2)} & & & \\ \vdots & & \ddots & \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_{784}^{(m)} \end{bmatrix}$$

(column vector is (784,1))

shape

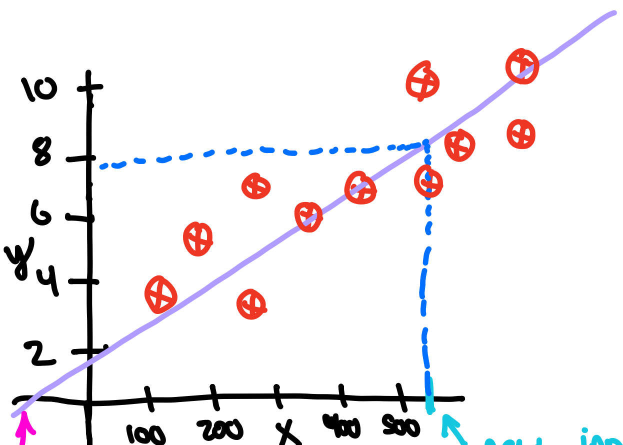
$y^{(i)} = \underline{2}$

no subscript

$$y^{(i)} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \leftarrow \text{"0"} \\ \leftarrow \text{"1"} \\ \leftarrow \text{"2"} \\ \vdots \\ \leftarrow \text{"9"} \end{array}$$

One-hot encoding

Linear Regression



What is the relationship between input "x" and output "y"

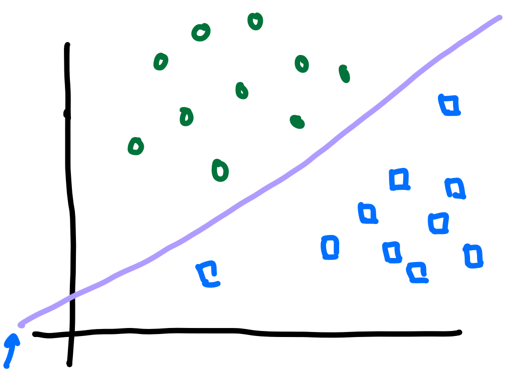
What is this line?

new input value, we don't know the true value for y

Model: $\hat{y} = mx + b$

$m = \frac{1}{50}$, $b = 1.9$

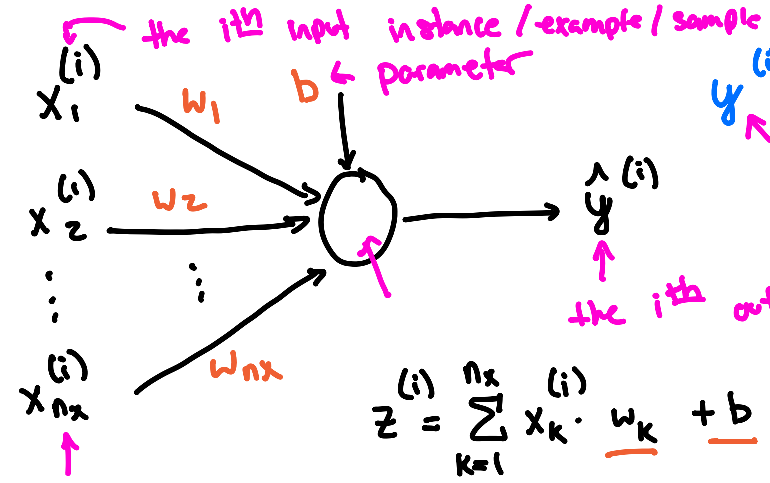
Binary Classification



$\hat{y} = mx + b > 0$

We want the output to be T/F, 1/0, 1/-1

Neuron



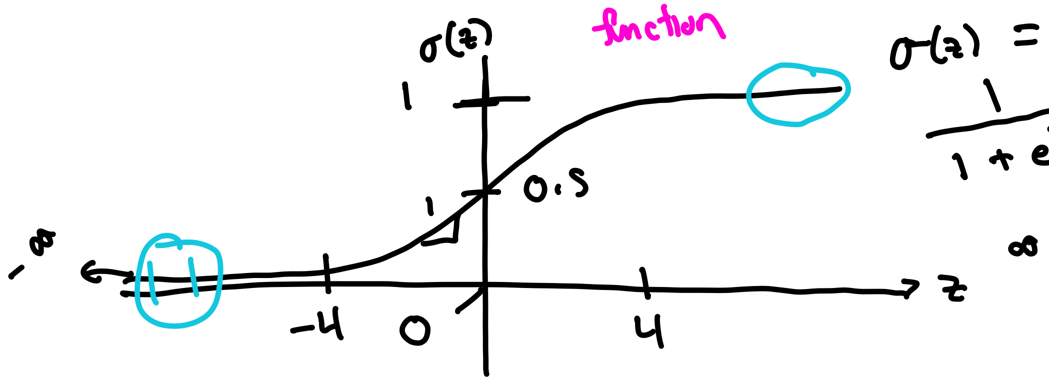
$y^{(i)}$ ← known label / target

$$z^{(i)} = \sum_{k=1}^{n_x} x_k^{(i)} \cdot \underline{w_k} + \underline{b} \quad \text{(linear)}$$

$$\hat{y}^{(i)} = \sigma\left(z^{(i)}\right) \quad \text{(activation)}$$

non-linear

activation / squashing function



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

Goal: We want to find values for w_k and b such that $\hat{y}^{(i)} \approx y^{(i)}$ for all values of i.

What is a good $\underbrace{\text{objective function}}_{\text{loss}}$?

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

Mean Difference

$y - \hat{y}$

Does not work very well.

$\hat{y}$	y	$\mathcal{L}$
0	0	0
0	1	-1
1	0	1
1	1	0

This is a problem

$$\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = |\hat{y}^{(i)} - y^{(i)}|$$ MAD