# A Single Neuron

- ✓ Data
- ✓ Regression vs. Classification
- Neuron

$$X \longrightarrow \underset{\theta,\ W+b}{\boxed{\text{Model}}} \longrightarrow \hat{y}$$

(guess)

## Data

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

input, label (known)

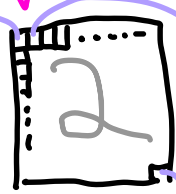
Supervised Learning

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

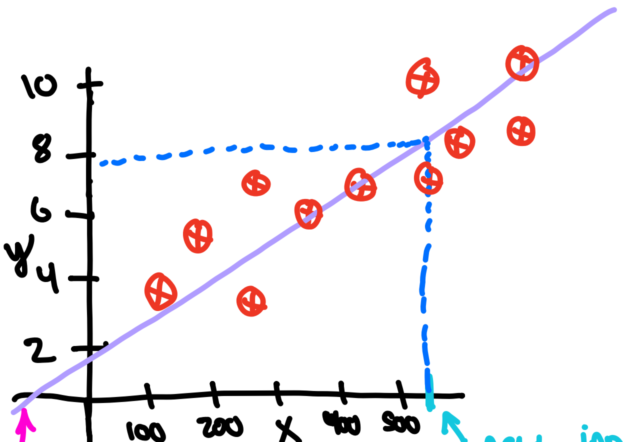
$y^{(i)} = \underline{2}$

no subscript

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

One-hot encoding

Linear Regression



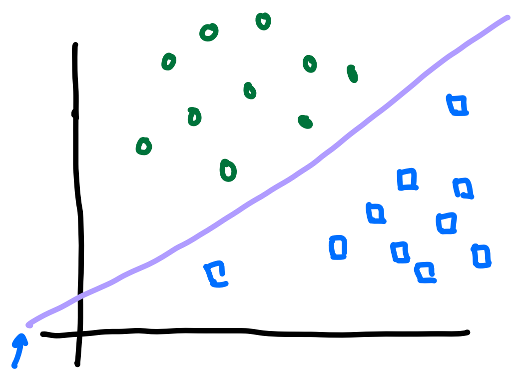

What is the relationship between input "x" and output "y"

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

What is this line?

Model: $\hat{y} = \underset{\frac{1}{50}}{m}x + \underset{1.9}{b}$

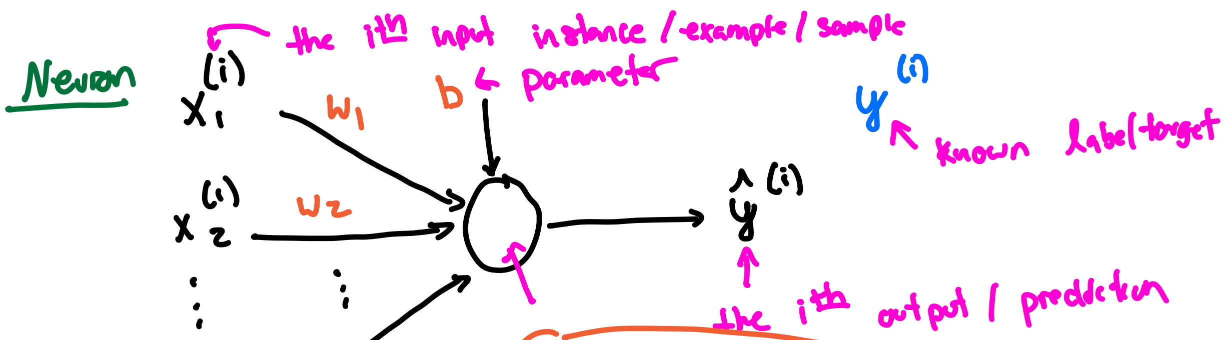
Binary Classifier



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

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

## Neuron

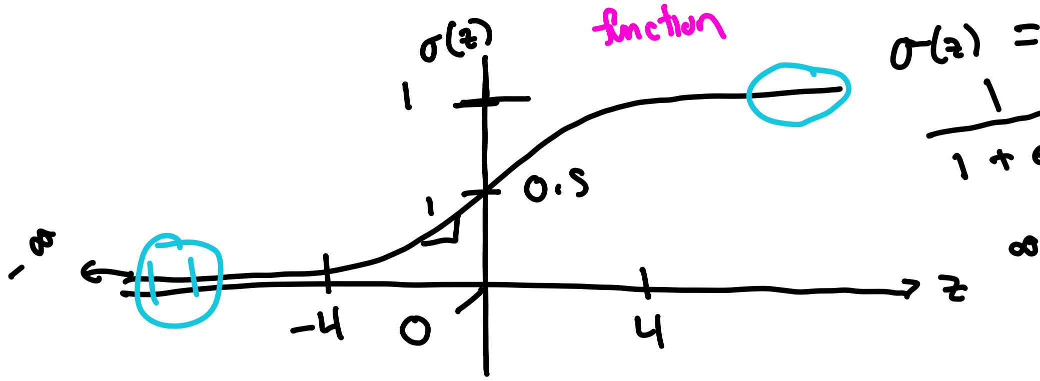



$$z^{(i)} = \sum_{k=1}^{n_x} x_k^{(i)} \cdot w_k + b$$ (linear)

# of input features

$$\hat{y}^{(i)} = \sigma(z^{(i)})$$ (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 $\underbrace{\text{objective function}}_{\text{loss}}$?

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

$$\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = \frac{1}{N}\left|\hat{y}^{(i)} - y^{(i)}\right| \quad \text{MAE}$$

$$\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = \frac{1}{N}(\hat{y}^{(i)} - y^{(i)})^2 \quad \text{MSE}$$

$$= (\hat{y}^{(i)} - y^{(i)})^2 \quad \text{SSE}$$

$$= \frac{1}{2N}(\hat{y}^{(i)} - y^{(i)})^2 \quad \text{Half-MSE}$$

$$\mathcal{L}(\hat{Y}, Y) = \frac{1}{2N}\sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)})^2 \quad \text{Equiv.}$$

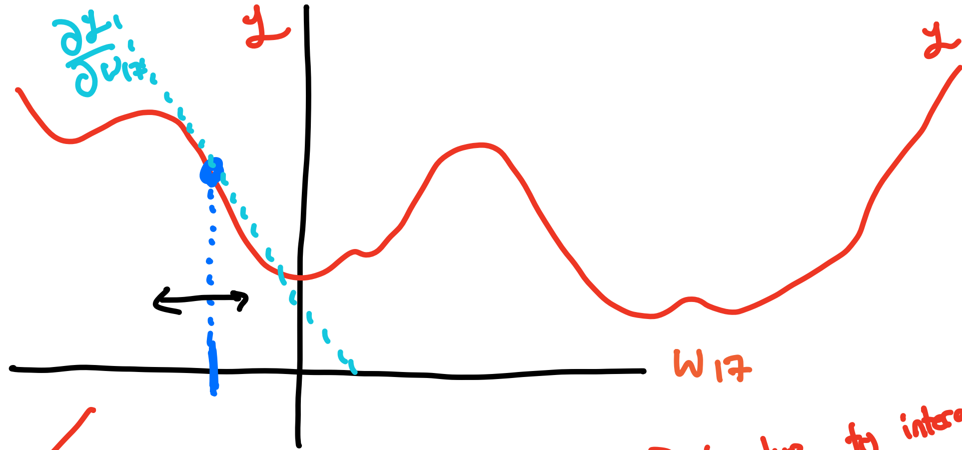
$$\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})$$

Binary Cross Entropy Loss

$\hat{y}^{(i)} \in [0,1]$ Range

$y^{(i)} \in \{0,1\}$ Set

How do I use $\mathcal{L}$ to find better values for $w_i$, $b$?




$\cancel{\frac{\partial \mathcal{L}}{\partial w_{17}} = 0}$ Closed-Form solution

Bad due to interactions among parameters

$\frac{\partial \mathcal{L}}{\partial w_{17}}$

$\partial$ partial derivative

$d$ derivative

$$w_{17} = w_{17} - \alpha \frac{\partial \mathcal{L}}{\partial w_{17}}$$

learning Rate

$$\frac{\partial \mathcal{L}}{\partial w_k} = \frac{\partial \; y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log(1-\hat{y}^{(i)})}{\partial w_k}$$

$$= \frac{\partial \mathcal{L}^{(i)}}{\partial \hat{y}^{(i)}} \cdot \frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} \cdot \frac{\partial z^{(i)}}{\partial w_k}$$

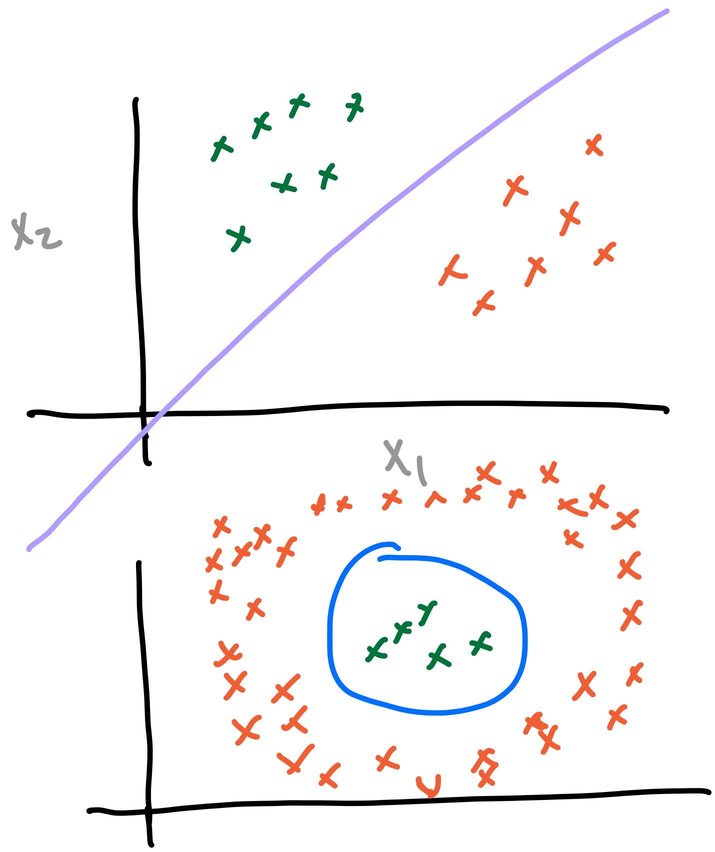
$$= \frac{1}{N} \sum_{i=1}^{N} (\hat{y}^{(i)} - y^{(i)}) X$$

$$\frac{\partial \mathcal{L}}{\partial b} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}^{(i)} - y^{(i)})$$

loop

$$w_k = w_k - \alpha \frac{\partial \mathcal{L}}{w_k}$$

$$b = b - \alpha \frac{\partial \mathcal{L}}{b}$$




★ Single Neurons cannot handle data that is not $\underbrace{\text{linearly}}_{\text{hyperplanerly}}$ separable.

→ Neural Network