

DL consultation

BOTOS CSABI



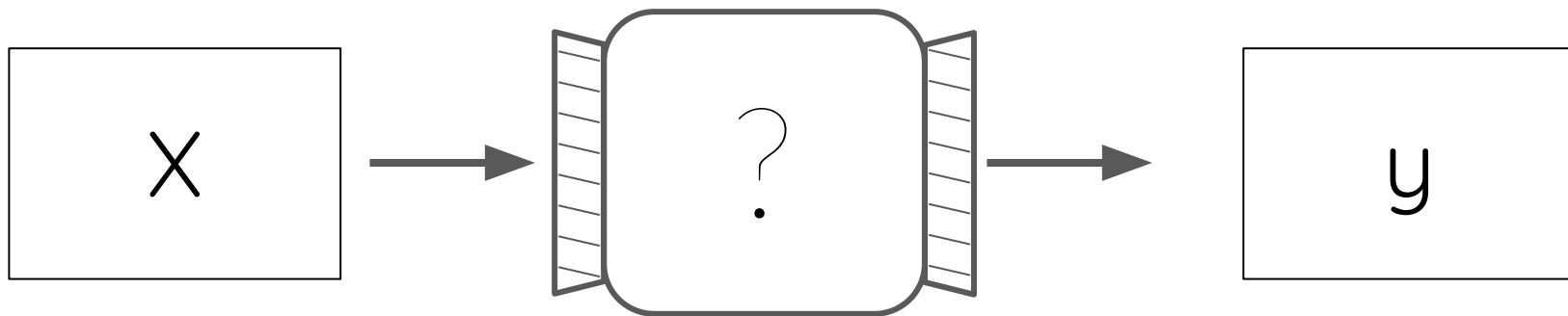
Today's topics

1. Programming or Teaching
2. Parametric functions
3. Distance
4. Optimization (intro)

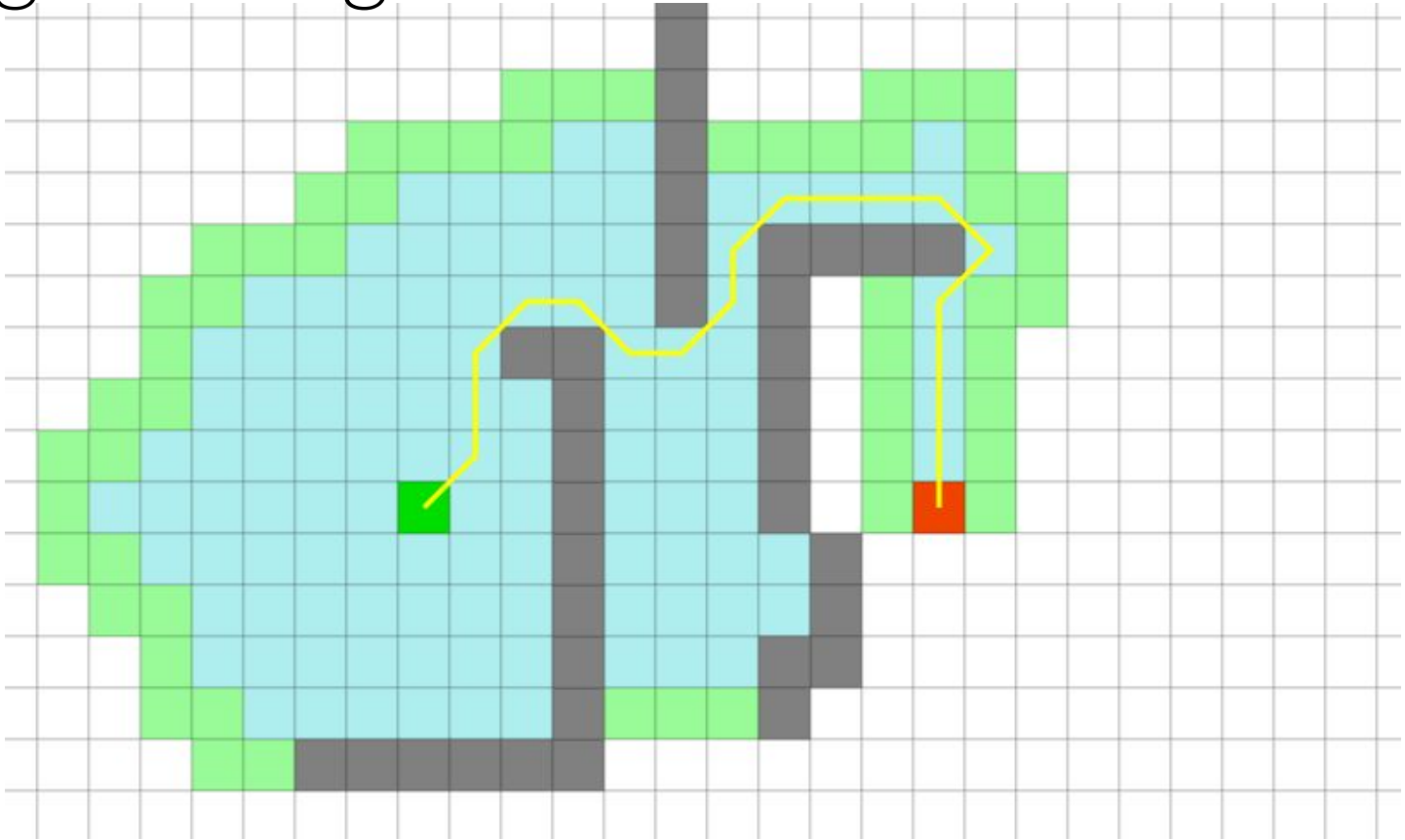
Who knows this guy?

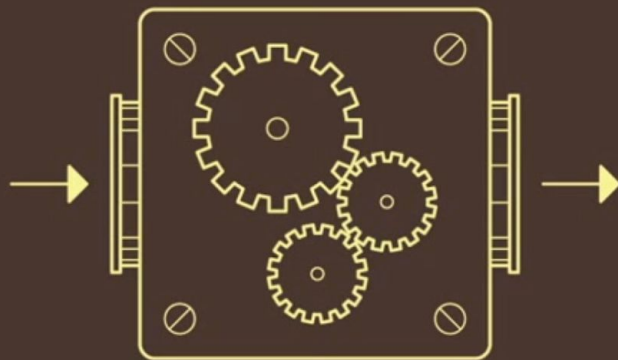


1. Programming or Teaching



Programming



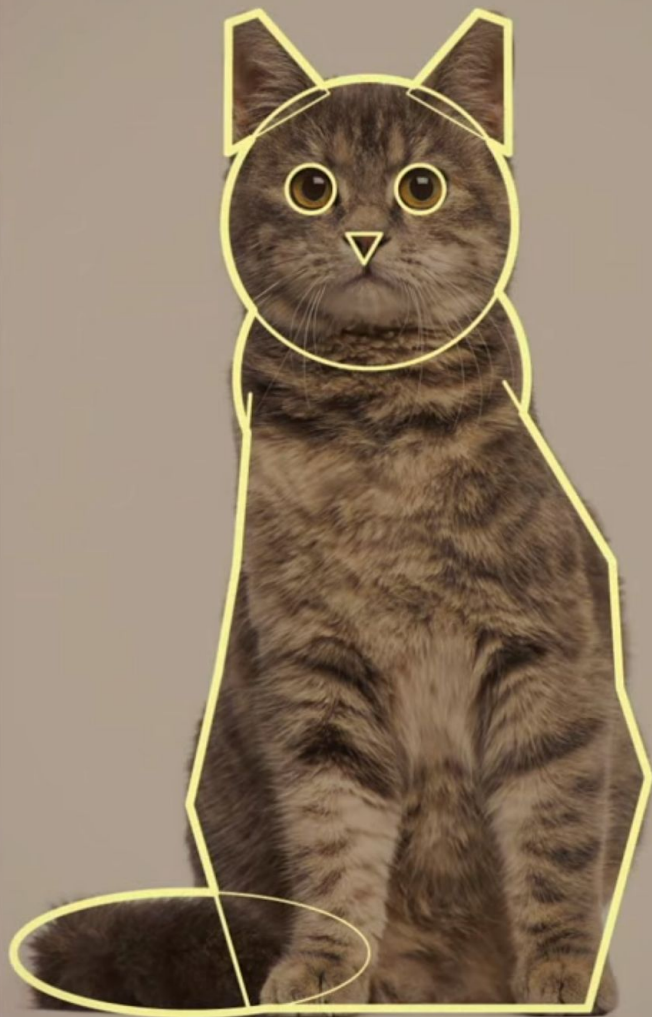


“Cat”

Credits: Fei Fei Li



Credits: Fei Fei Li



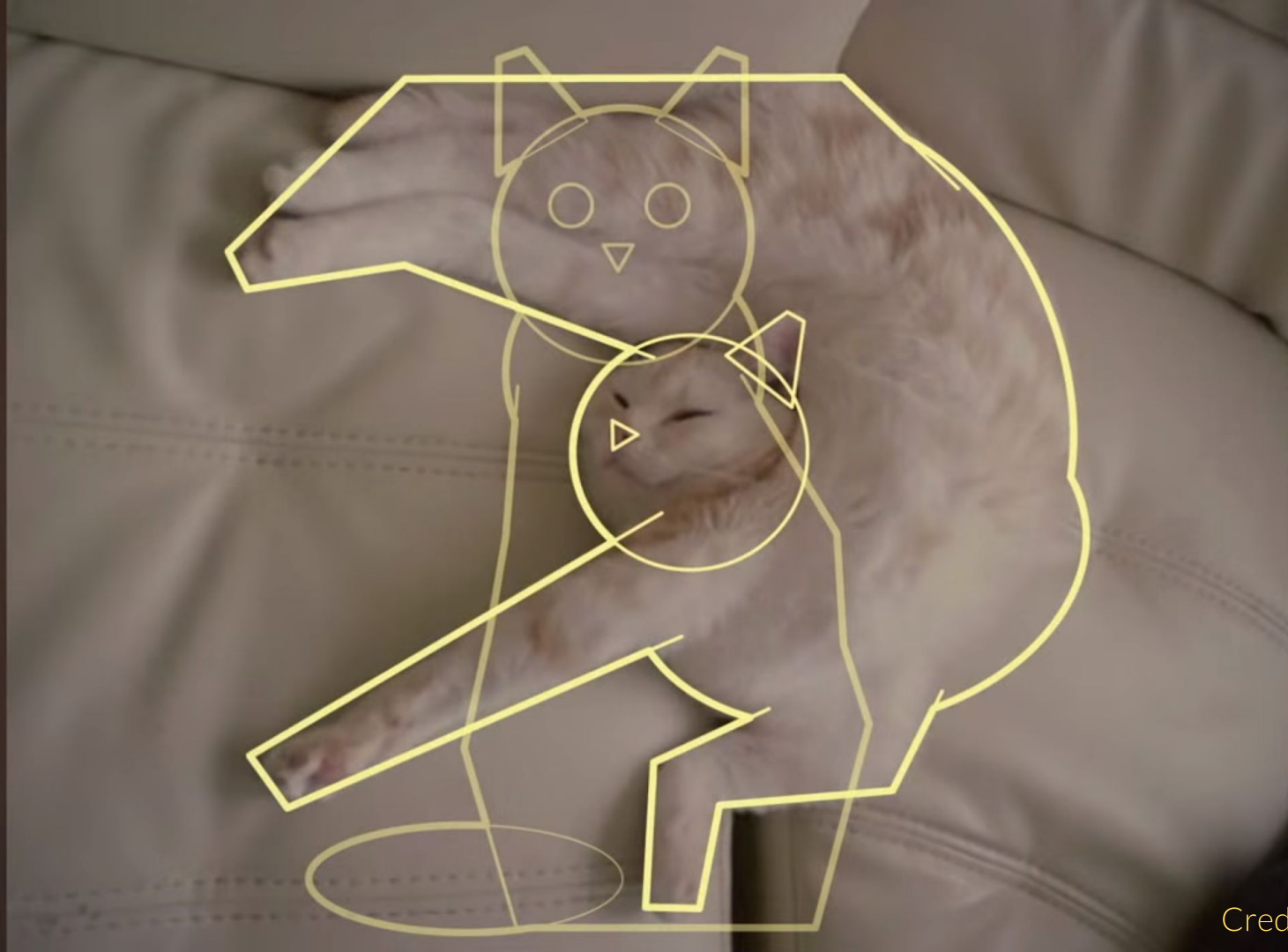
Credits: Fei Fei Li



“Cat”



Credits: Fei Fei Li



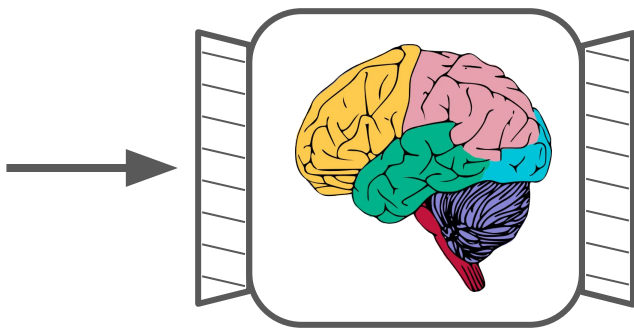
Credits: Fei Fei Li



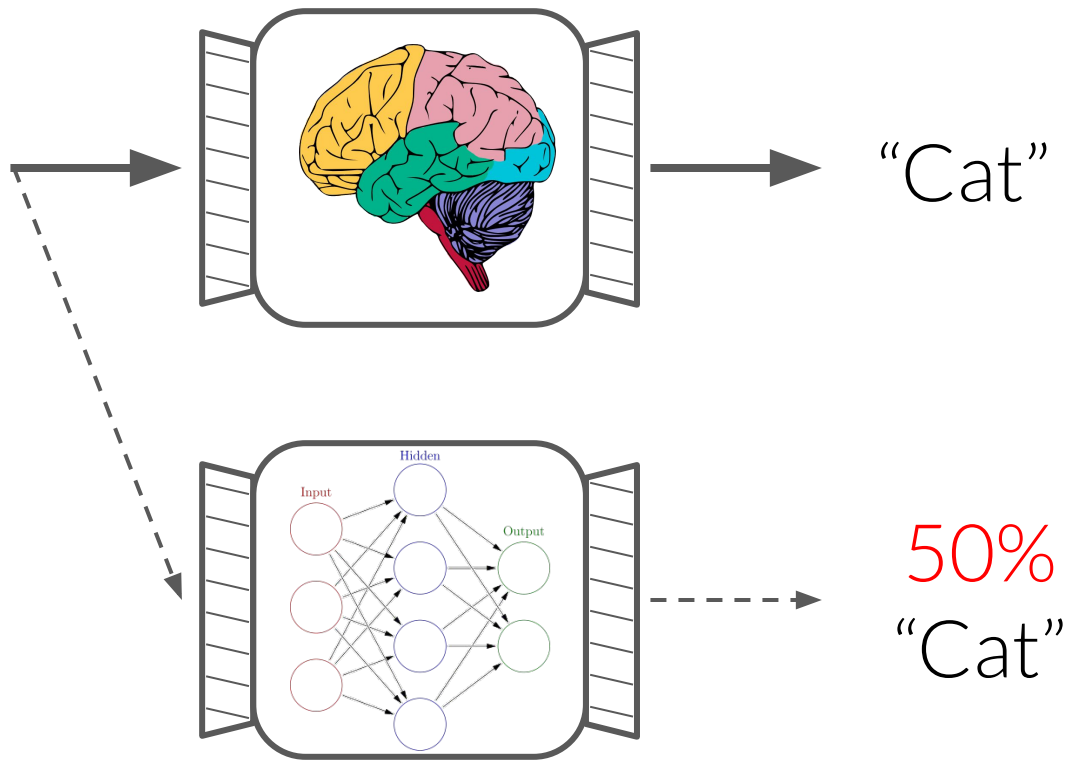
Credits: Fei Fei Li

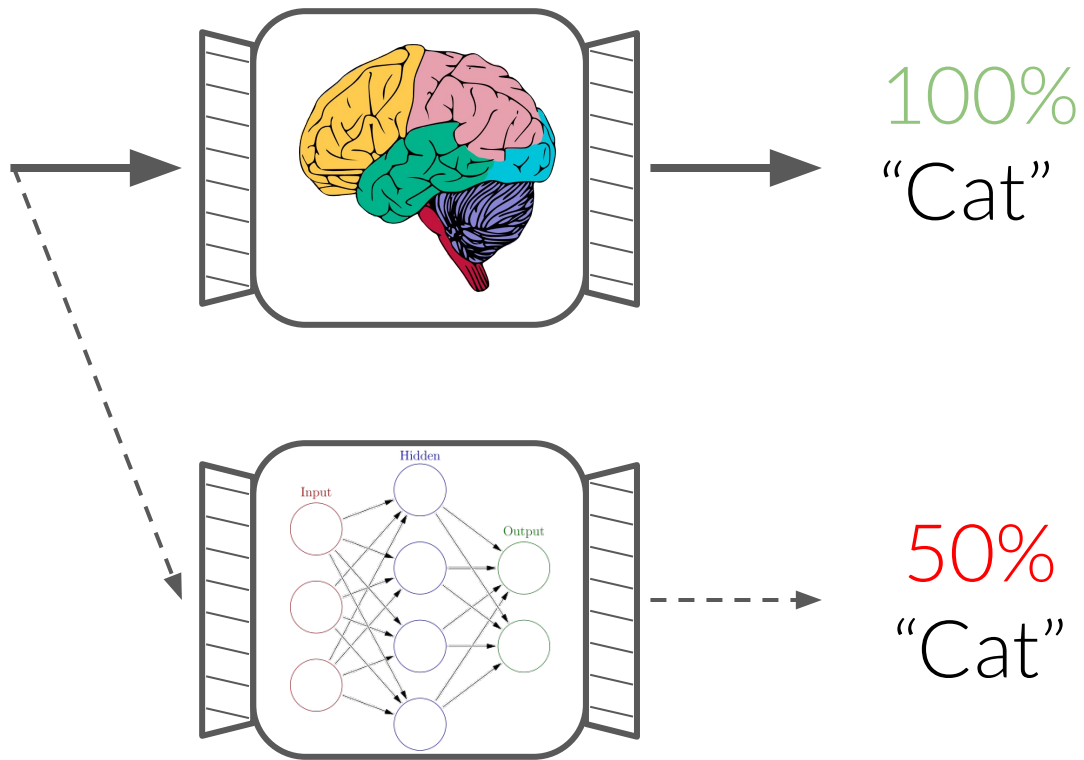
General truth: there are no general truths

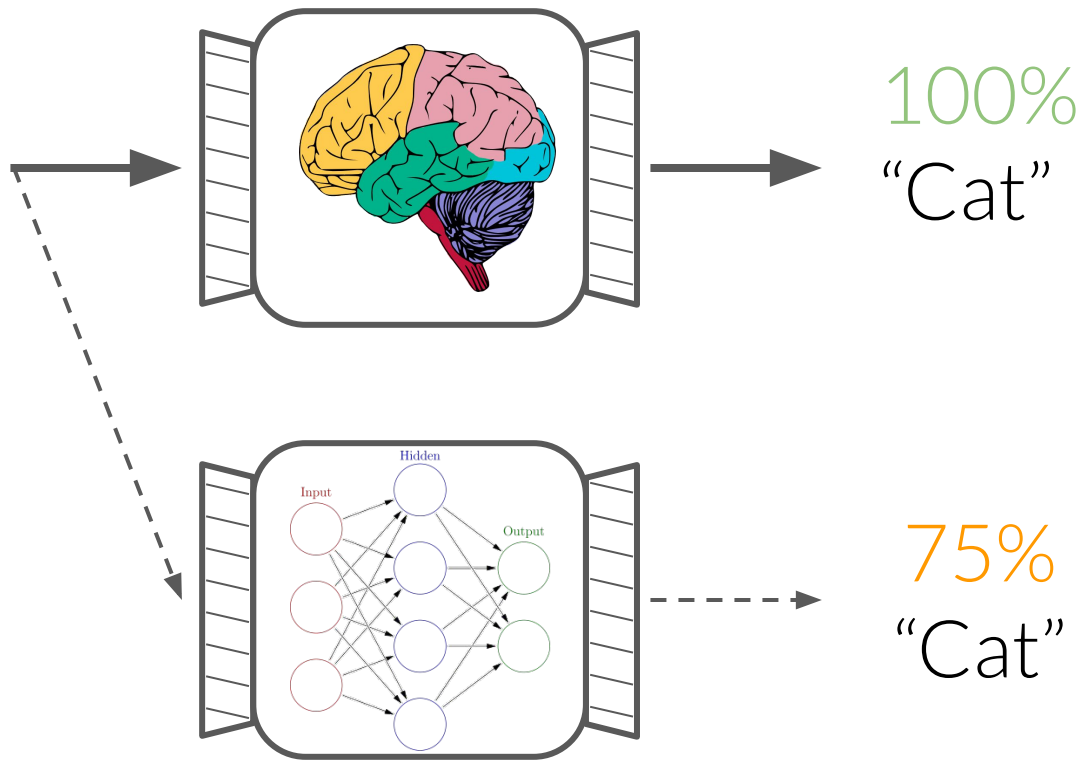


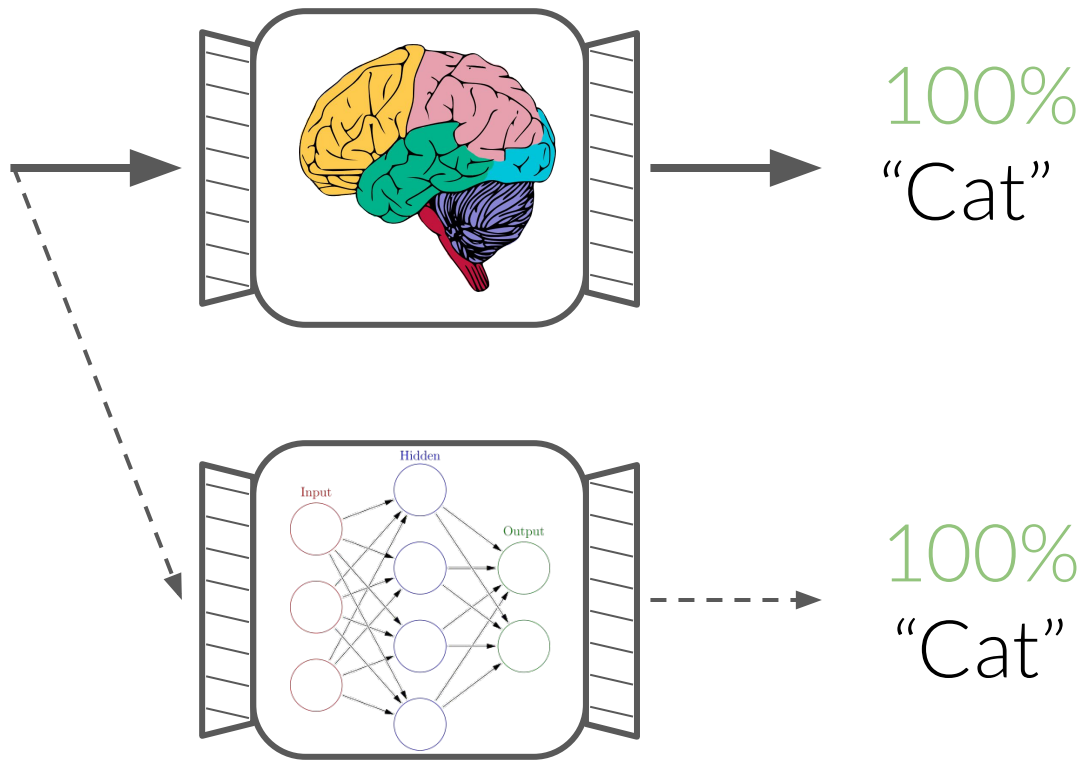


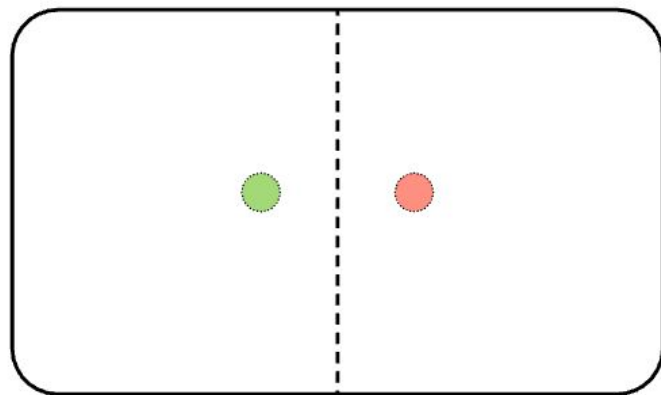
“Cat”

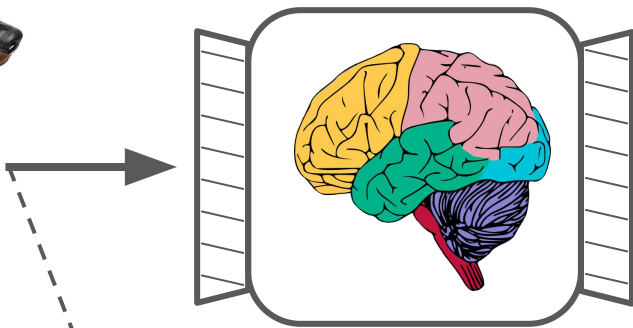




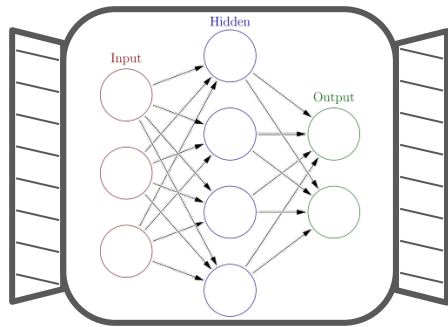
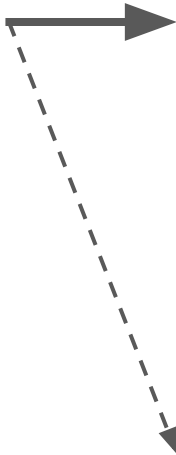






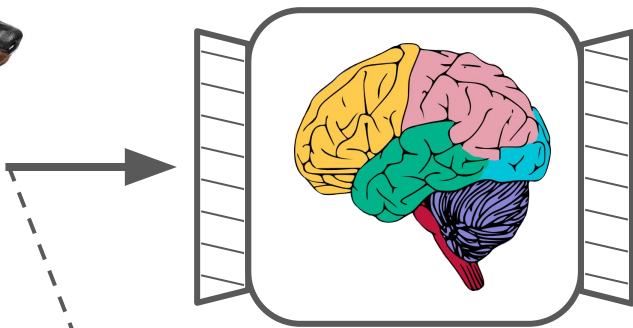


0%
“Cat”

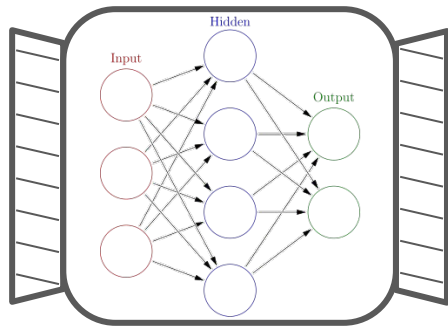
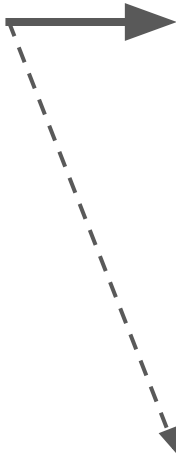


100%
“Cat”

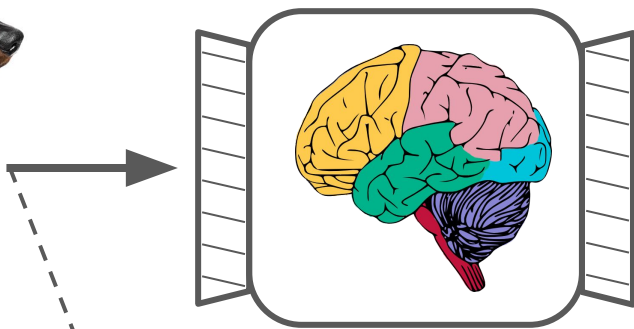




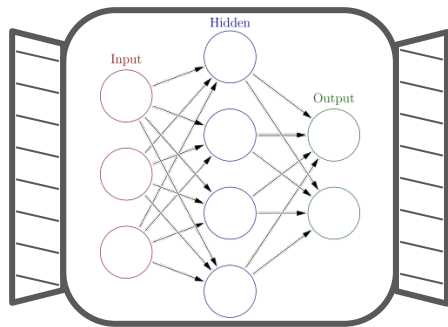
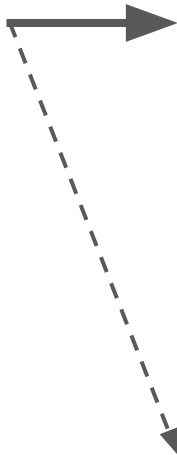
0%
“Cat”



50%
“Cat”

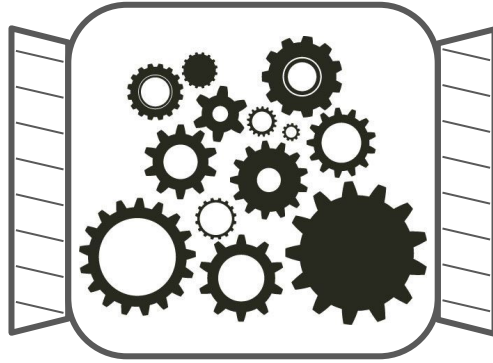


0%
“Cat”

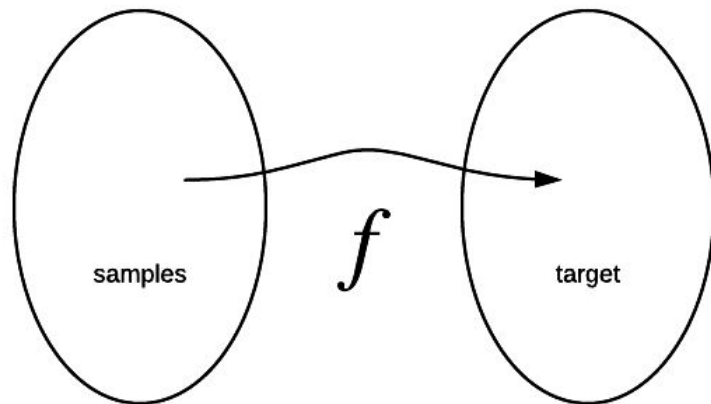


0%
“Cat”

2. Parametric functions



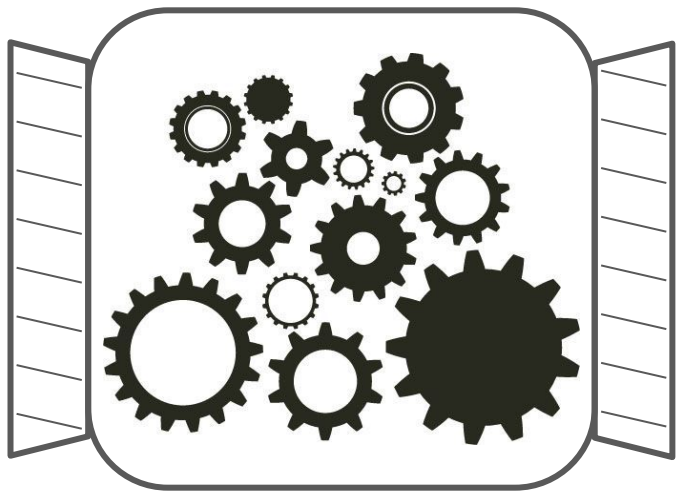
Parametric functions



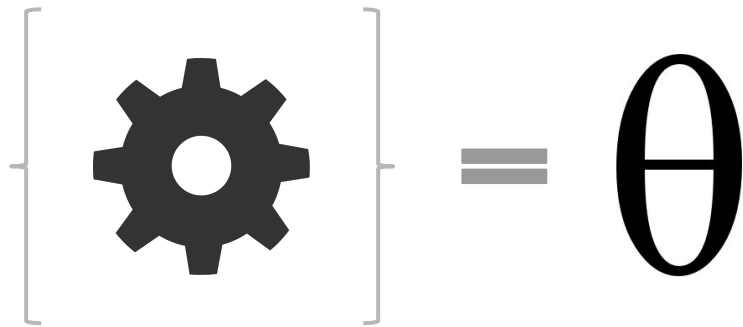
$$f : \mathbf{X} \rightarrow \mathbf{Y}$$

$$f(x) = f_{\theta}(x)$$

Parametric functions

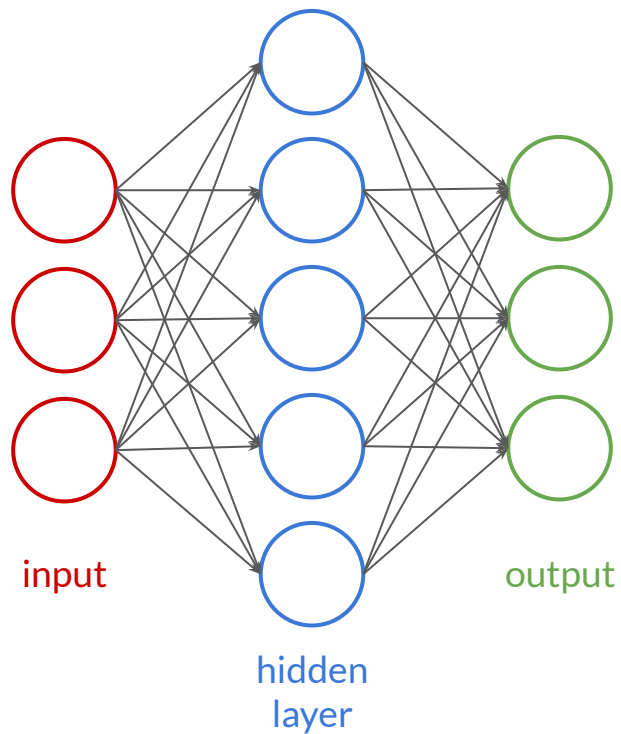


$$f : X \rightarrow Y$$

A diagram showing a single gear icon centered within a large left-facing curly bracket. To the right of the bracket is an equals sign, followed by the Greek letter θ .

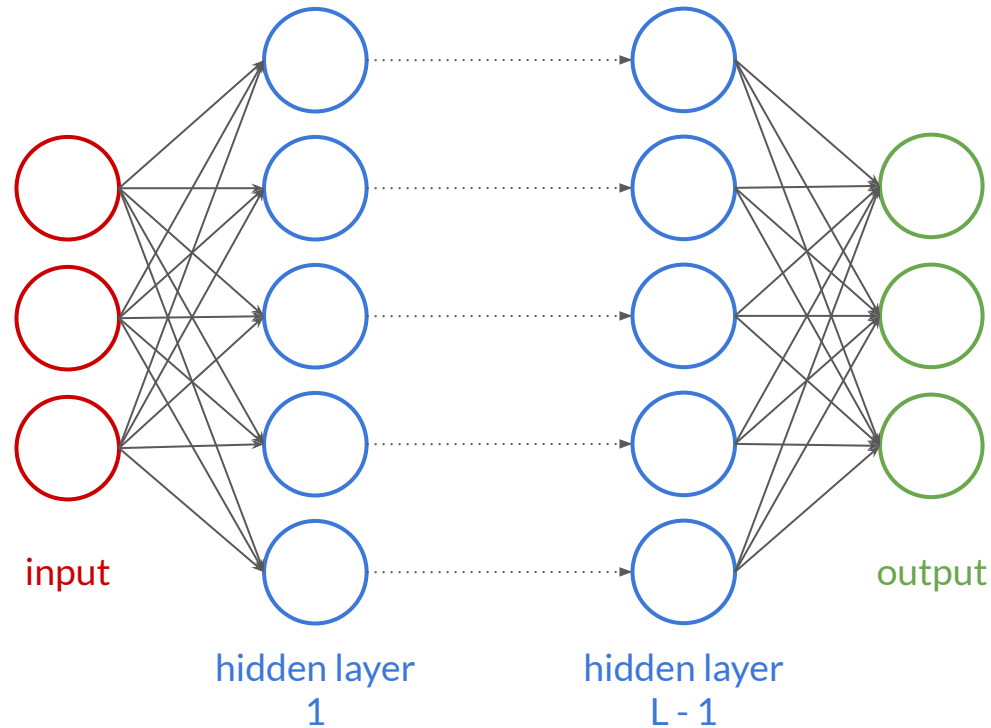
$$f(x) = f_{\theta}(x)$$

Composition

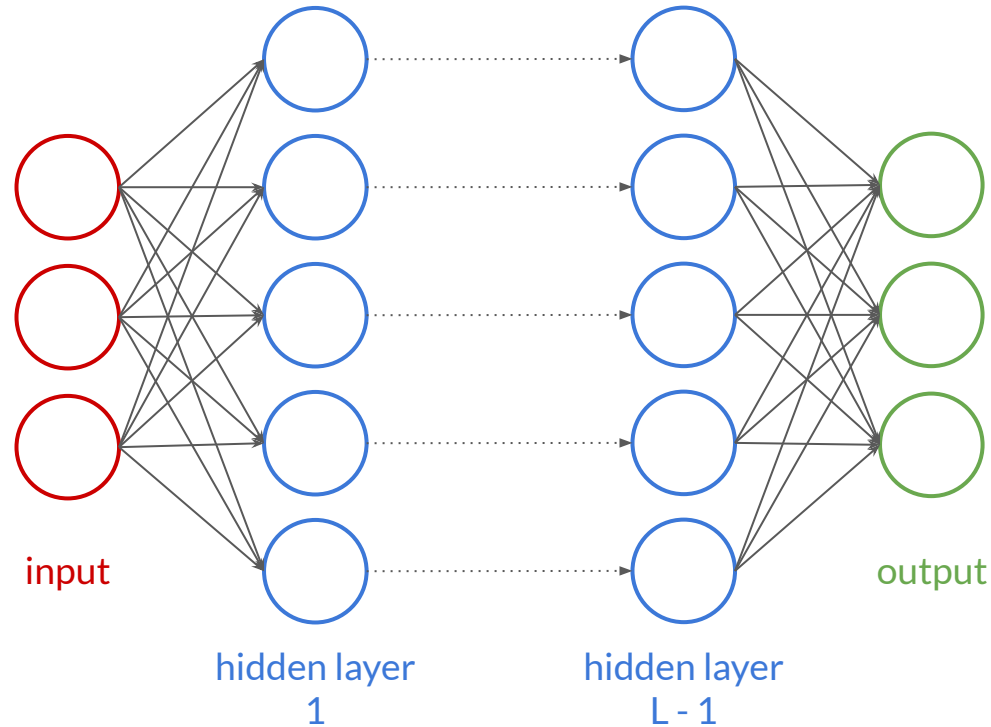


$$a2(a1(x))$$

$$a_L (a_{L-1} (a_{L-2} (\dots a_1 (x))))$$



$$a^L \circ a^{L-1} \circ a^{L-2} \circ \dots \circ a^1 (x)$$



$$a(x) = \text{non-lin} (\langle \mathbf{w}, \mathbf{x} \rangle + b)$$

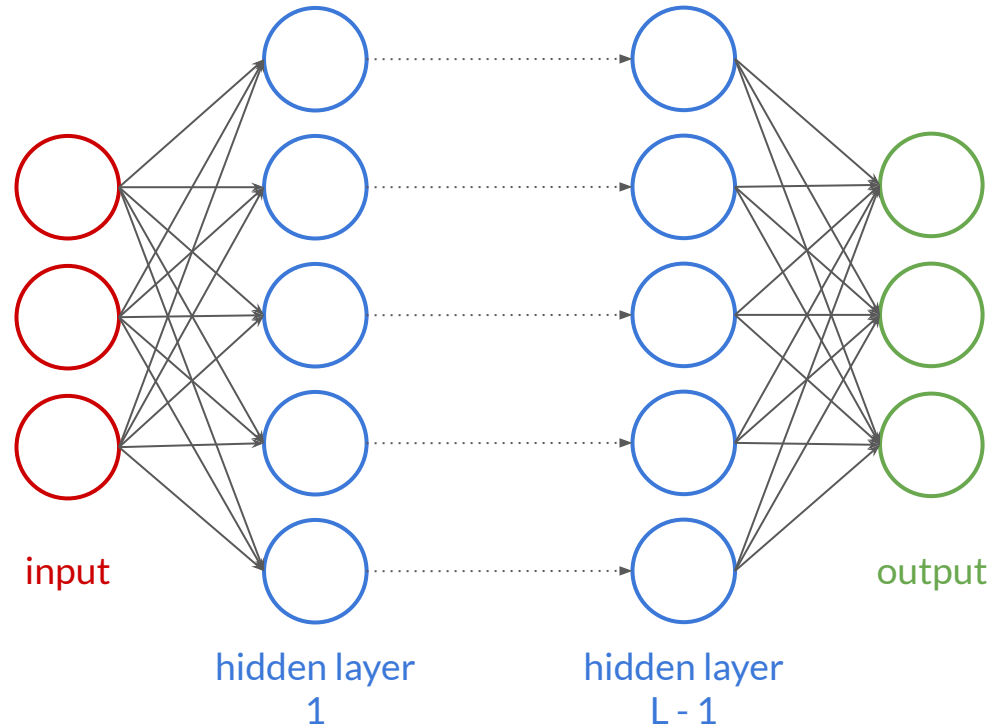
$$\mathbf{a}(x) = \text{non-lin} (\langle \mathbf{W}, \mathbf{x} \rangle + \mathbf{b})$$

$$\mathbf{a}^l(x) = \text{non-lin} (\langle \mathbf{W}, \mathbf{a}^{l-1}(x) \rangle + \mathbf{b})$$

$$a_j^l = \sigma \left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l \right)$$

$$\mathbf{a}^l(\mathbf{x}) = \text{non-lin} \left(\langle \mathbf{W}^l, \mathbf{a}^{l-1}(\mathbf{x}) \rangle + \mathbf{b}^l \right)$$

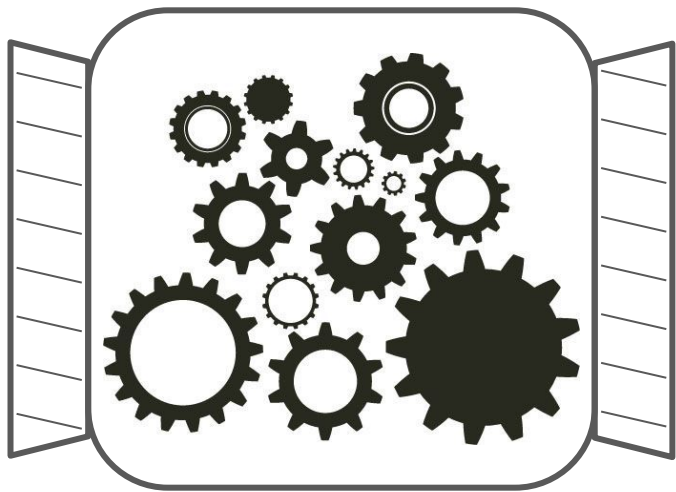
$$a^L \circ a^{L-1} \circ a^{L-2} \circ \dots \circ a^1 (x)$$



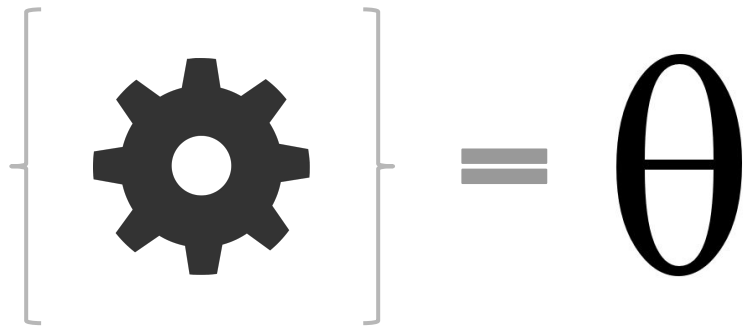
$$a_L \circ a_{L-1} \circ a_{L-2} \circ \dots \circ a_1(x)$$

```
a_val = x
for i in 1 ... L
    a_val = a_i(a_val)
```

Parametric function

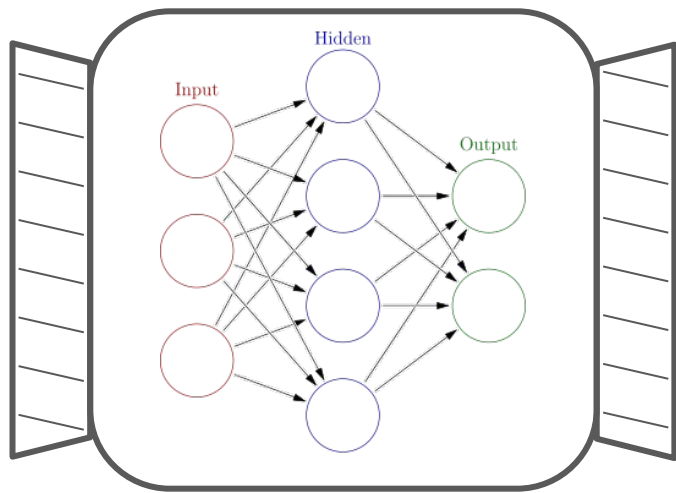


$$f : X \rightarrow Y$$

A diagram illustrating a parametric function. It features a single gear icon centered within a large, light gray left-facing curly bracket. To the right of the bracket is an equals sign, followed by the Greek letter θ . This visualizes the concept of a function defined by a single set of parameters.

$$f(x) = f_{\theta}(x)$$

Parametric function

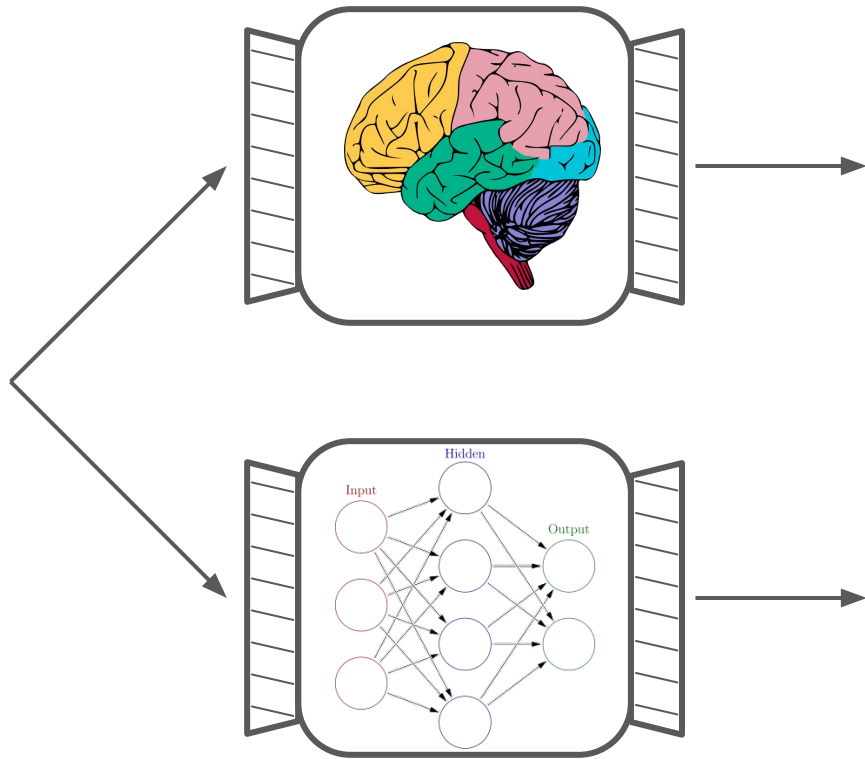


$$f : \mathbf{X} \rightarrow \mathbf{Y}$$

$$\{\mathbf{w}, \mathbf{b}\} = \boldsymbol{\theta}$$

$$f(x) = f_{\theta}(x)$$

X



y

d

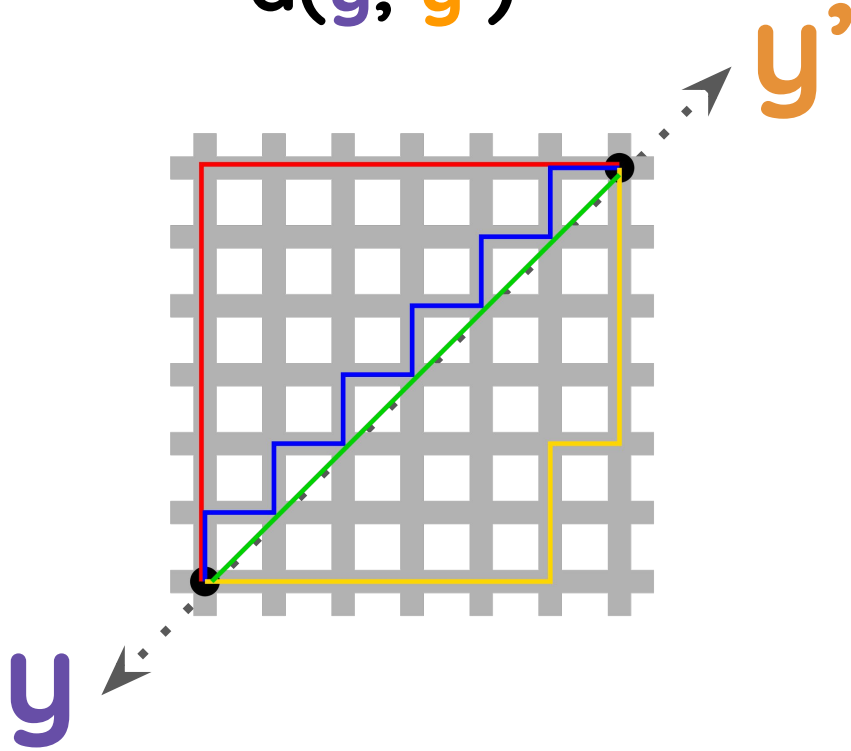
y'

3. Distance



$$y \xleftrightarrow{d(y, y')} y'$$

$$d(y, y')$$

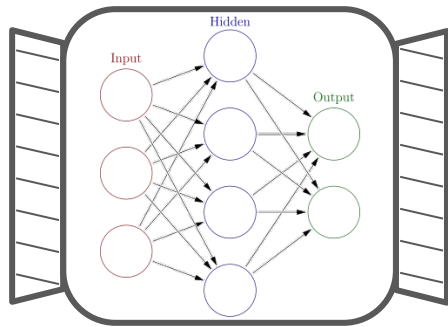


Classification example

Large distance



0%
“Cat”

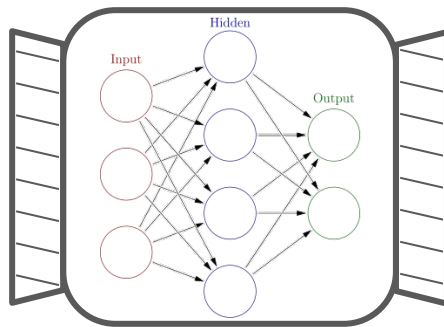


90%
“Cat”

Small distance

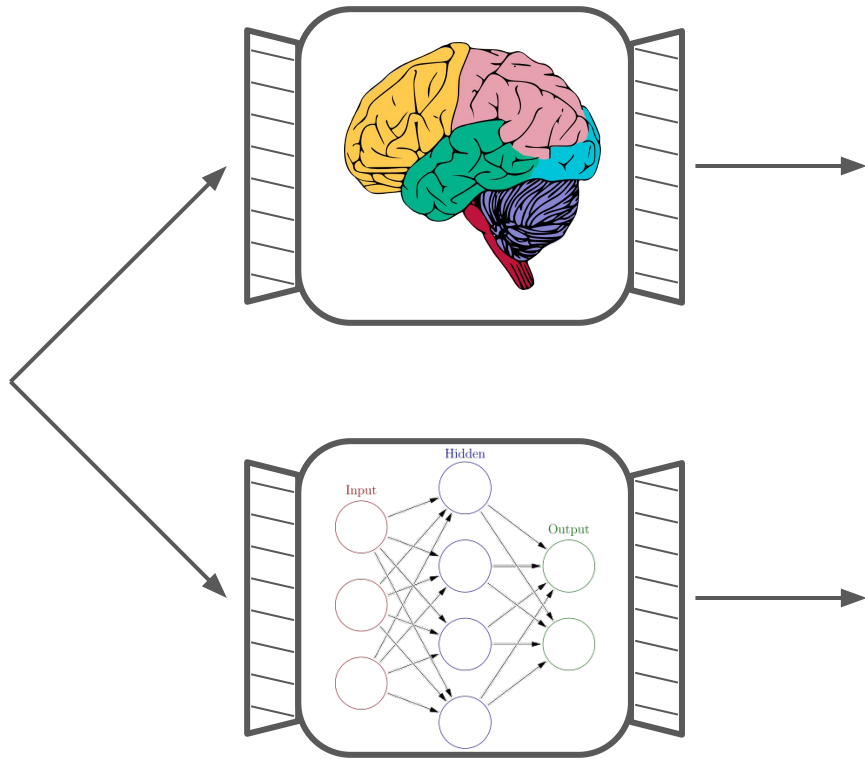


0%
“Cat”



3%
“Cat”

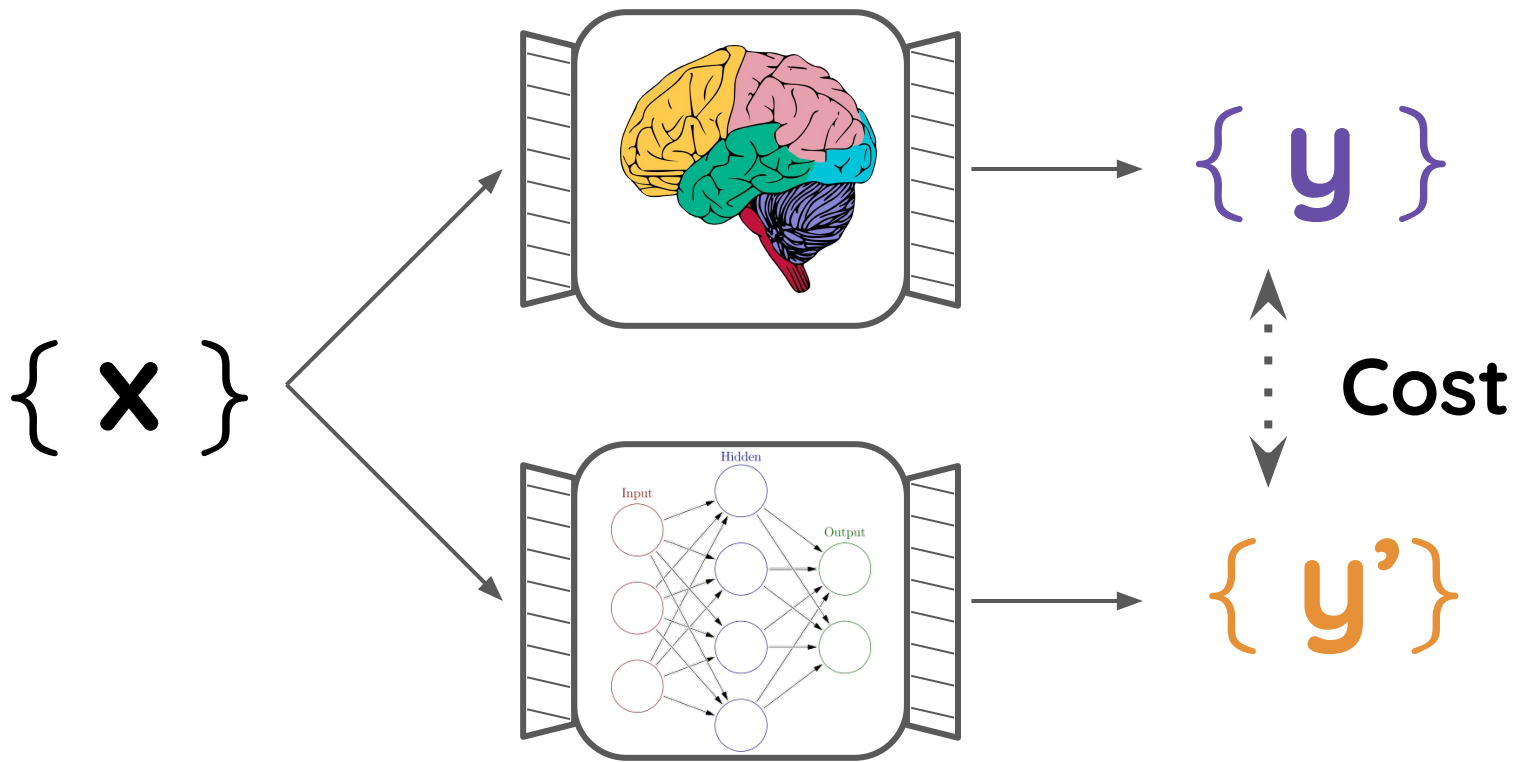
X



y

d

y'



Regression demo:

<https://goo.gl/3qJXKP>

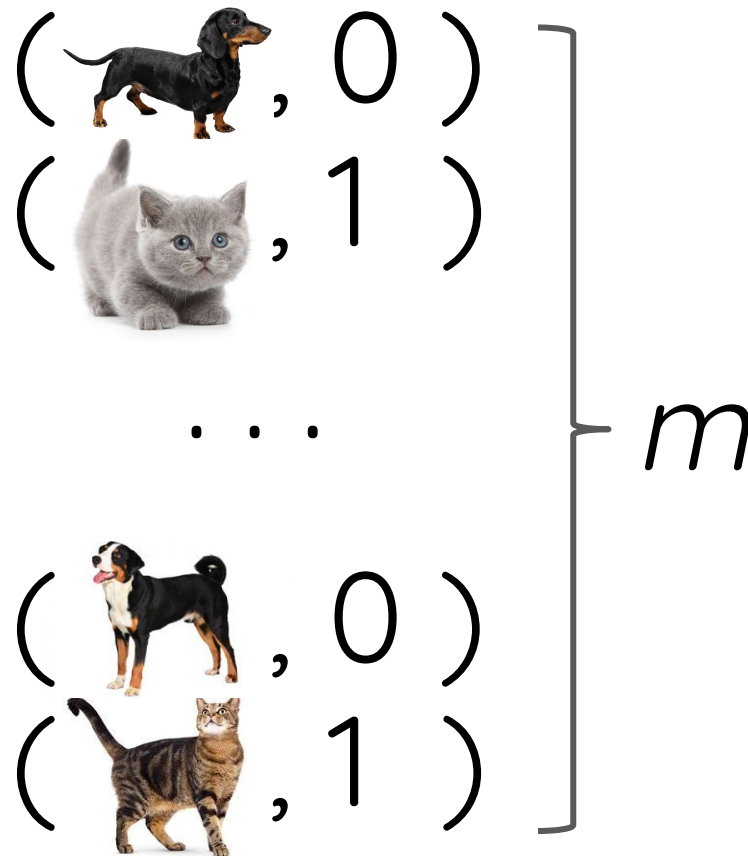
Cost function

Given m number of $(\mathbf{x}, y)$
sample pairs



Cost function

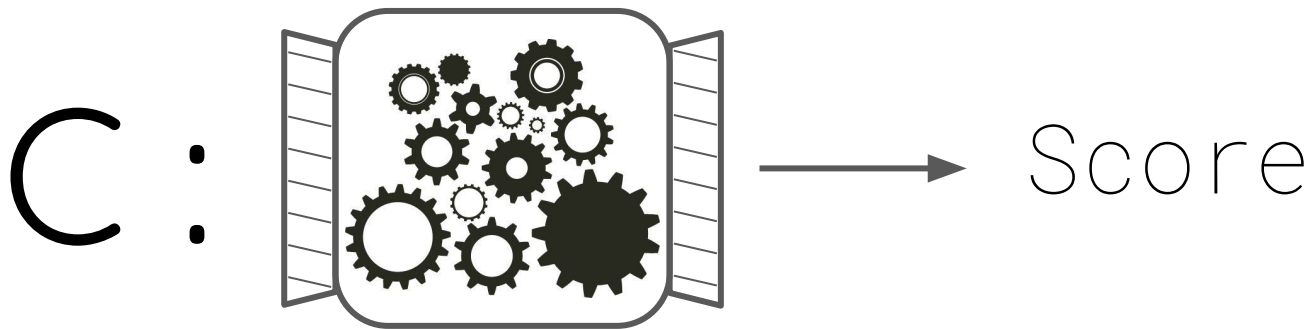
Given m number of $(\mathbf{x}, y)$
sample pairs



Cost function

Given m number of $(\mathbf{x}, \mathbf{y})$ sample pairs

The Cost function scores an *algorithm* that predicts $\mathbf{y}'$

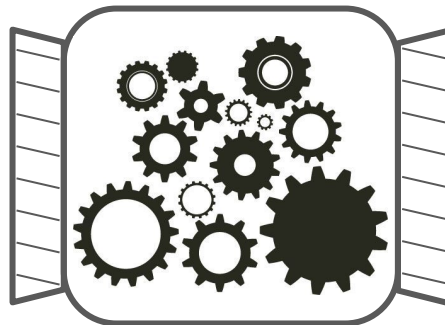


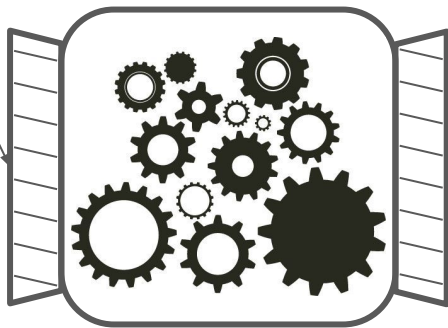


...

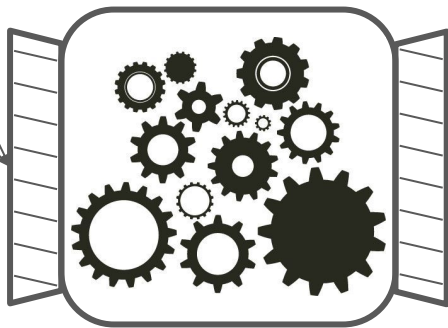


How does it work?





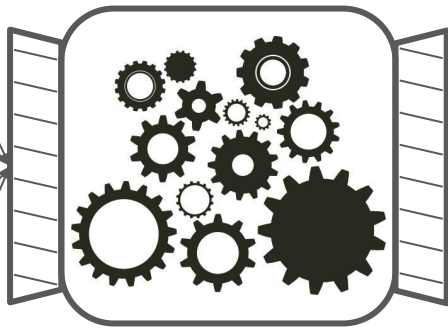
$(0.1, 0)$



$$\begin{pmatrix} 0.1, & 0 \end{pmatrix} \\ \begin{pmatrix} 0.8, & 1 \end{pmatrix}$$



...



$$\begin{pmatrix} 0.1, & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0.8, & 1 \end{pmatrix}$$

...

$$\begin{pmatrix} 0.4, & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0.3, & 1 \end{pmatrix}$$

$$\left[\begin{array}{cc} (0.1, 0) \\ (0.8, 1) \\ \vdots \\ (0.4, 0) \\ (0.3, 1) \end{array} \right]$$

$$(y', y)$$

$$\left[\begin{array}{cc} (0.1, 0) \\ (0.8, 1) \\ \dots \\ (0.4, 0) \\ (0.3, 1) \end{array} \right]$$

$$d(y', y)$$

(0.1, 0)

(0.8, 1)

...

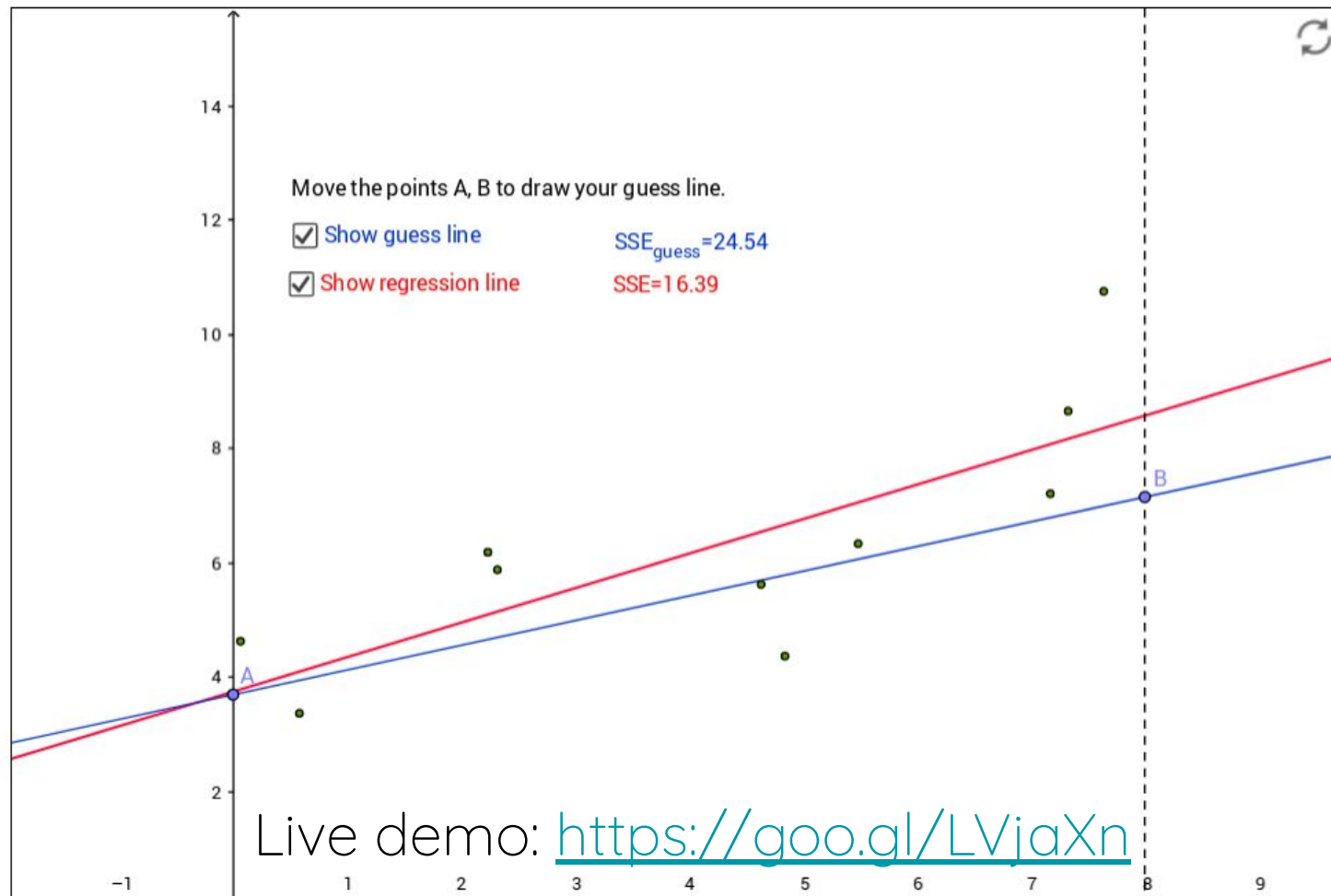
(0.4, 0)

(0.3, 1)

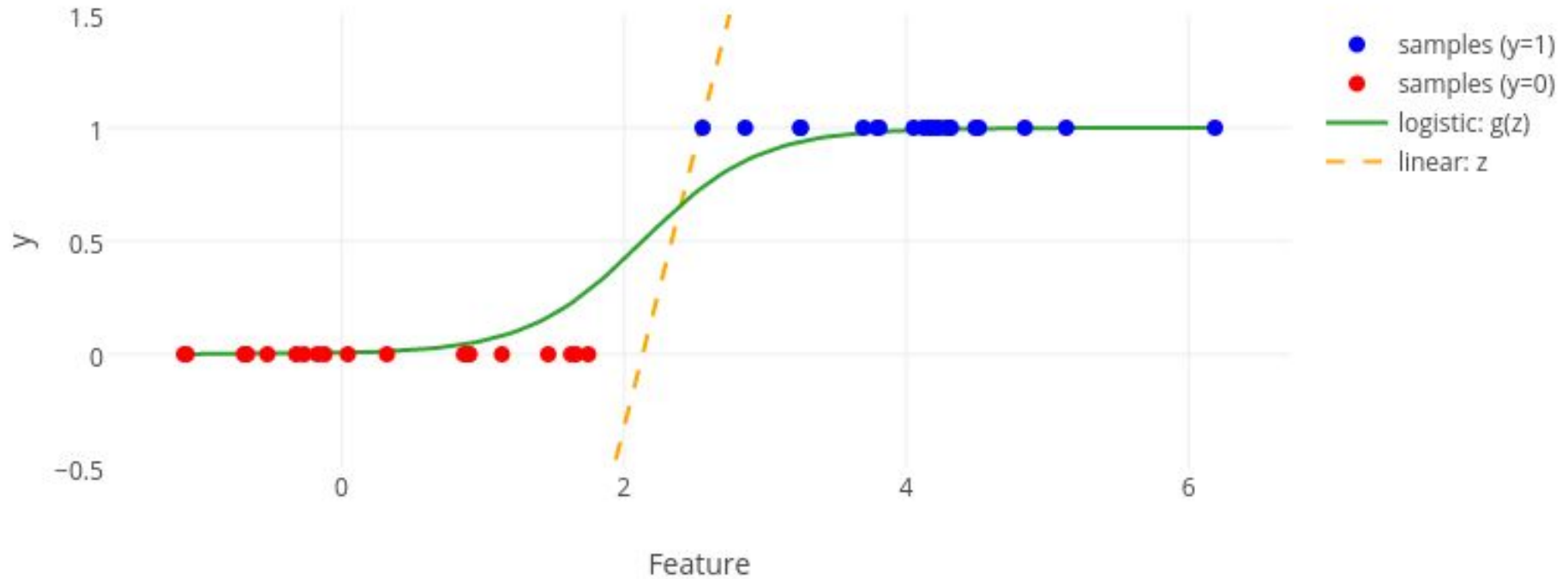
$$y' \xleftrightarrow[d(y', y)]{(y' - y)^2} y$$



Linear Regression Demo



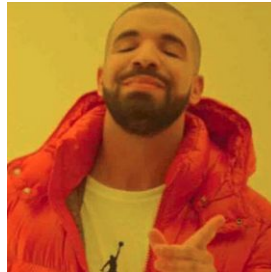
Logistic Regression: 1 Feature



Live demo: <https://goo.gl/UMQAHu>

$$y' \xleftrightarrow{d(y', y)} y$$

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

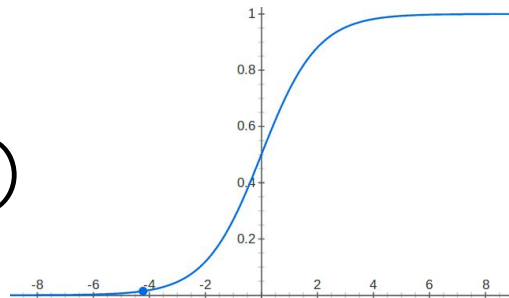


Binary Cross-entropy

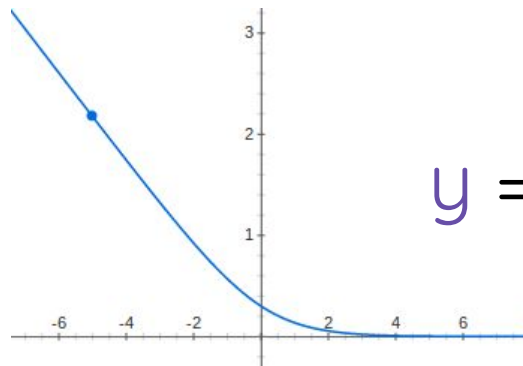
$$- y \log(y') - (1-y) \log(1 - y')$$

$$y' = 1 / (1 + \exp(-aL))$$

$$y' = 1 / (1 - \exp(-aL))$$

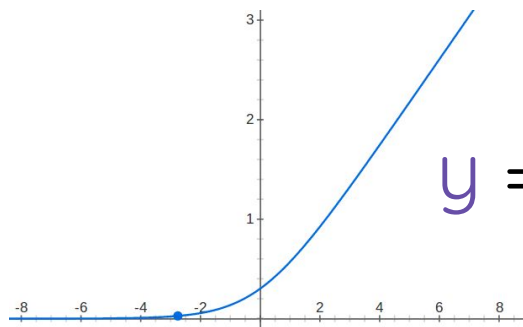


$$-y \log(y')$$



$$y = 1$$

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

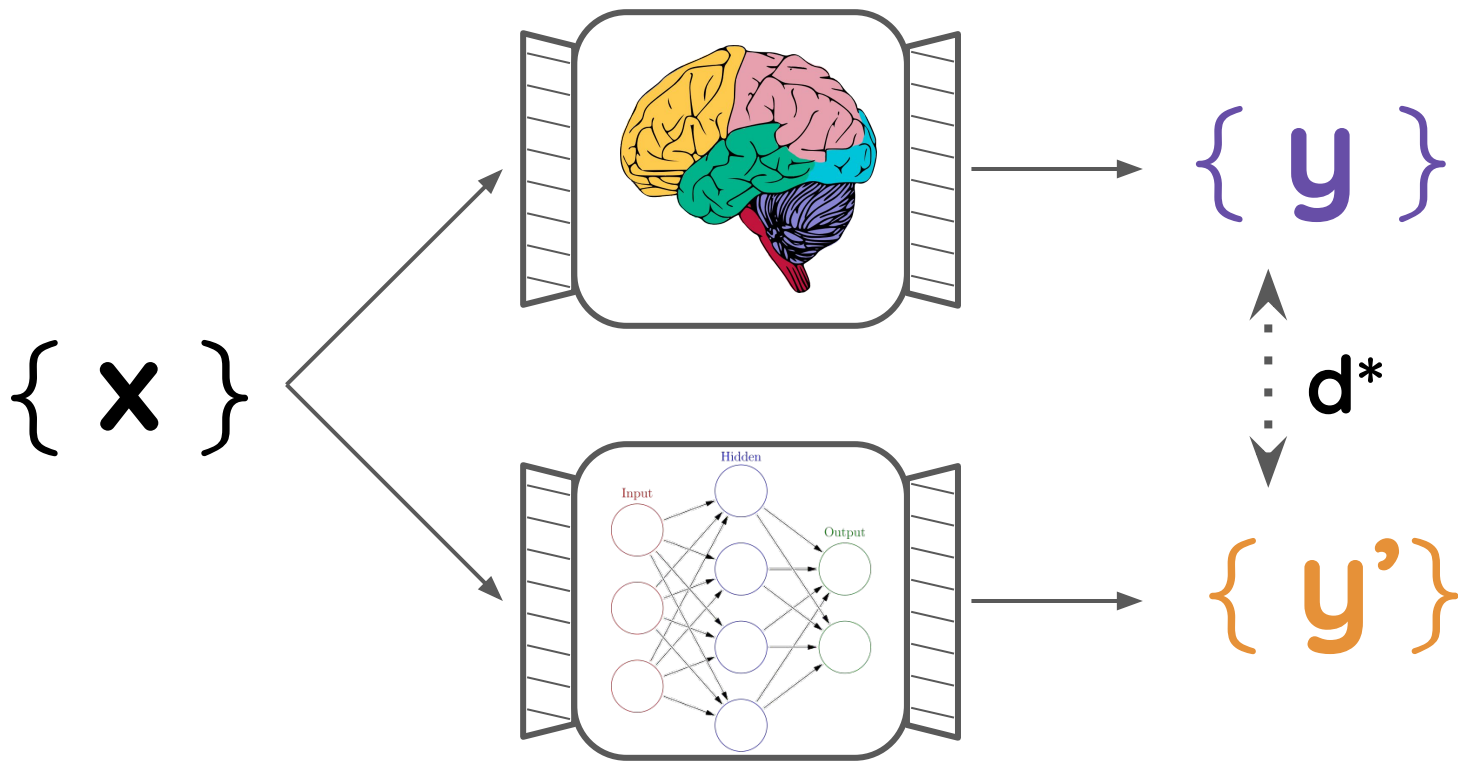


$$y = 0$$

Binary Cross-entropy

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

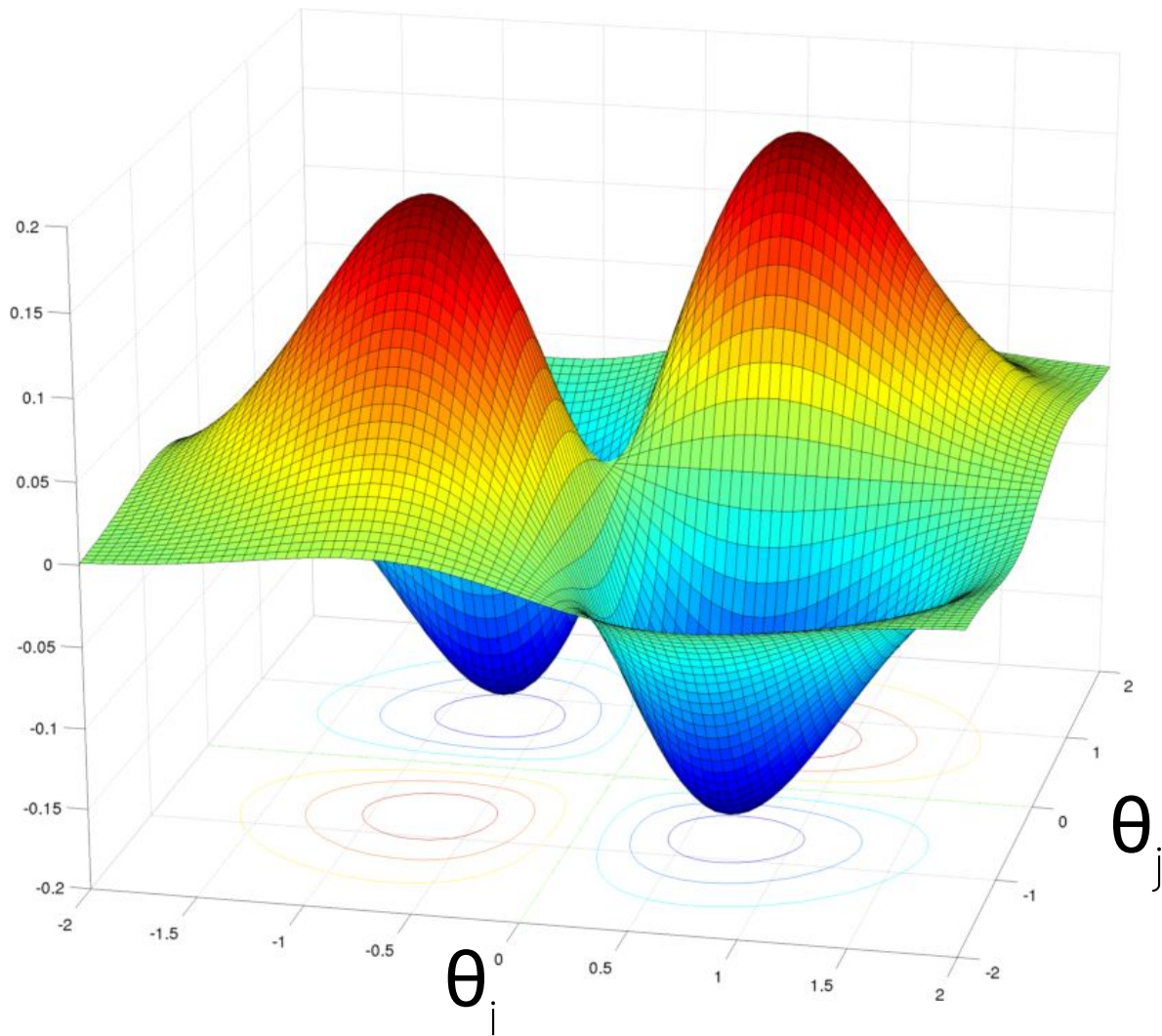
$$y' = 1 / (1 + \exp(-aL))$$



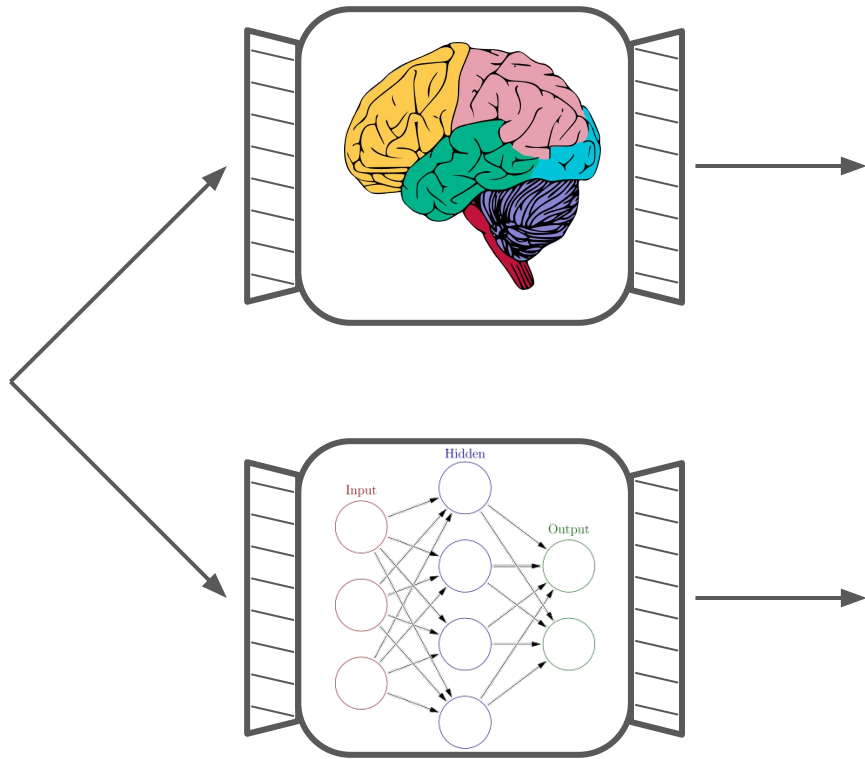
Binary Cross-entropy Cost

$$\sum - y \log(y') - (1-y) \log(1 - y')$$

$C(\theta)$



X

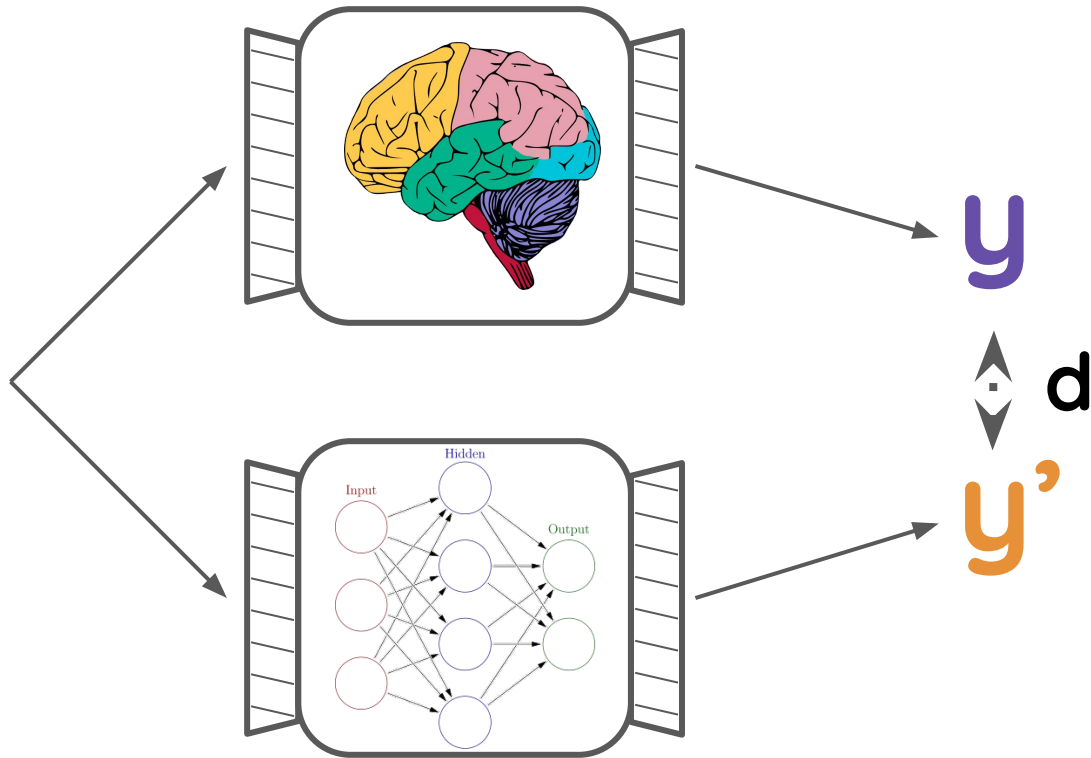


y

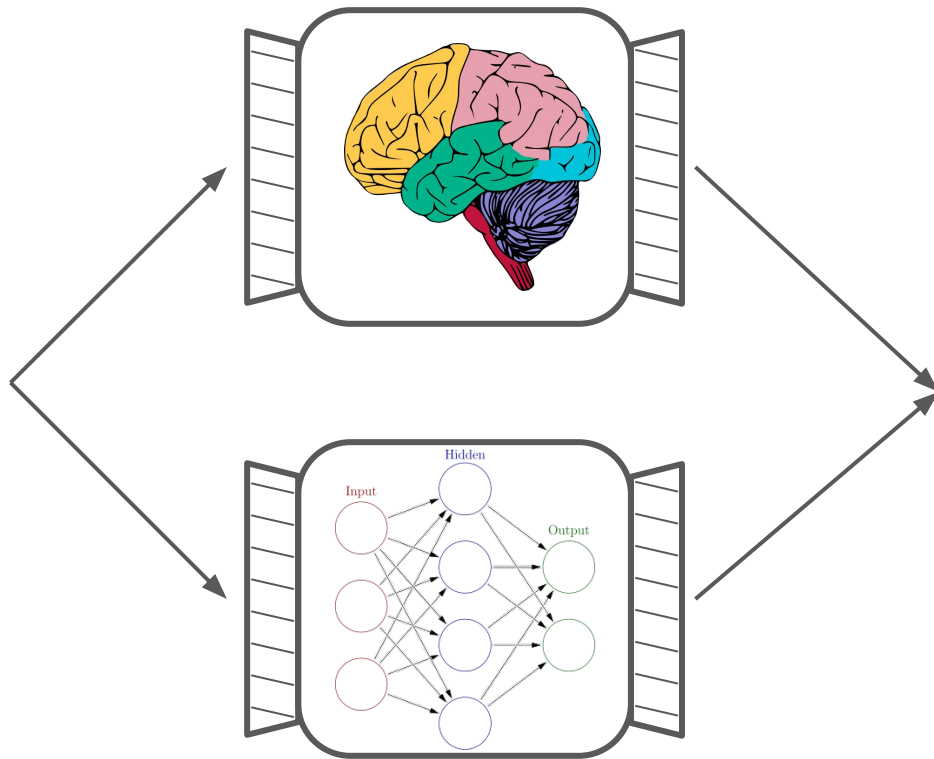
α

y'

X



x

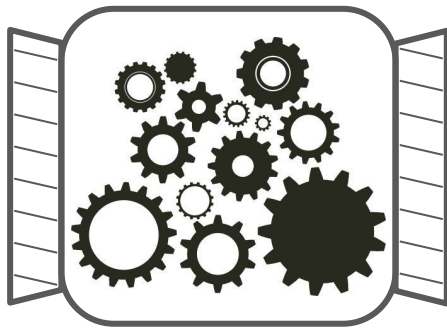


$$y = y'$$

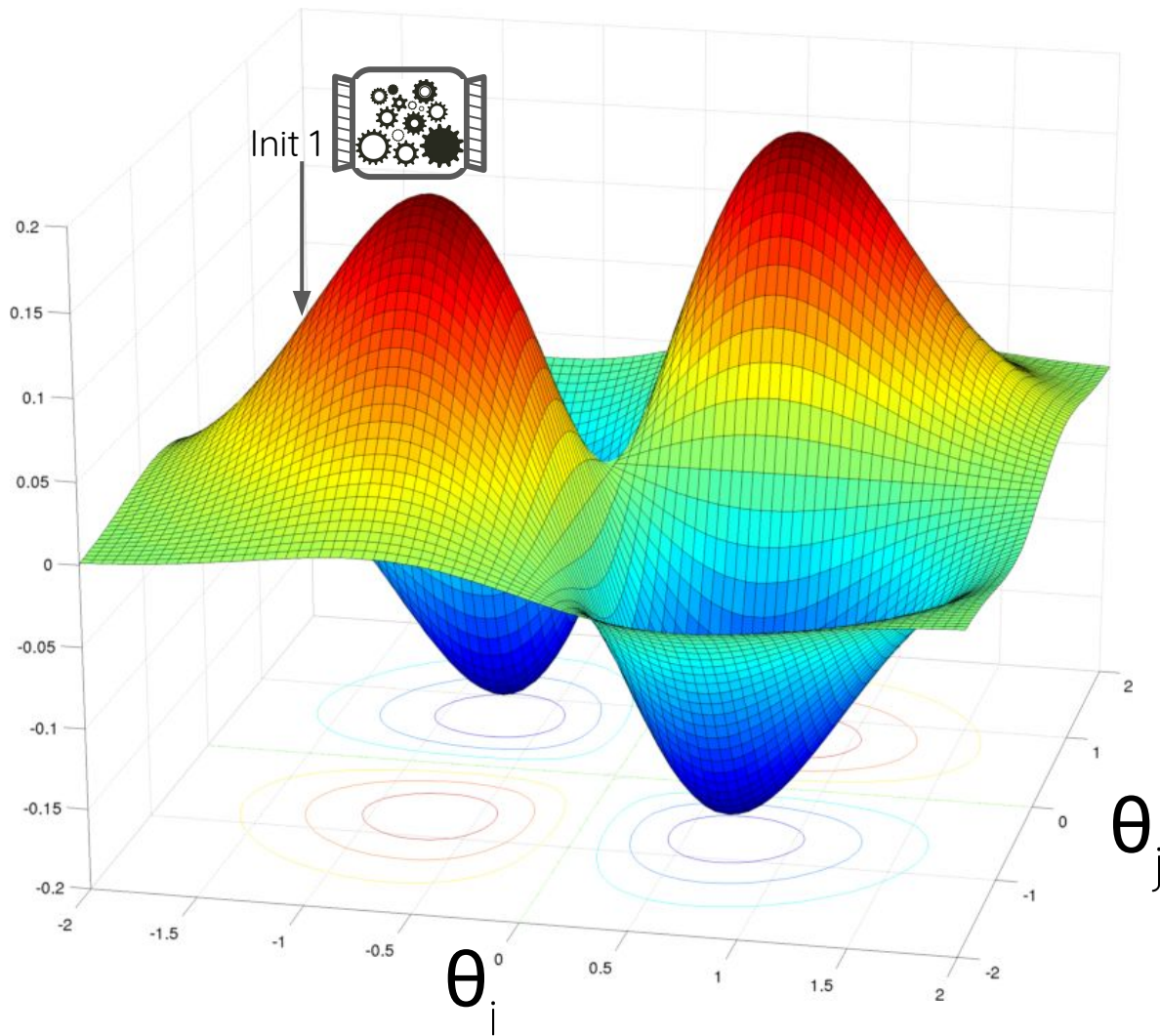
4. Optimization



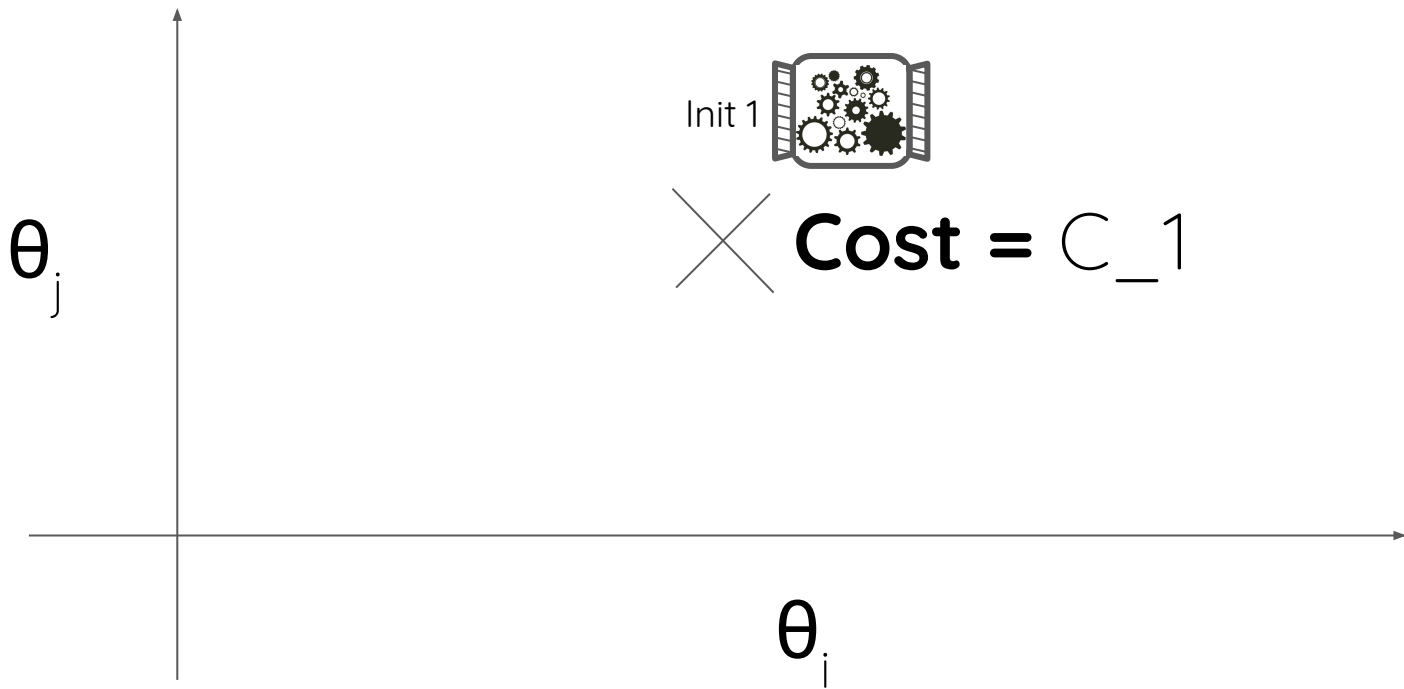
Tuning the parameters



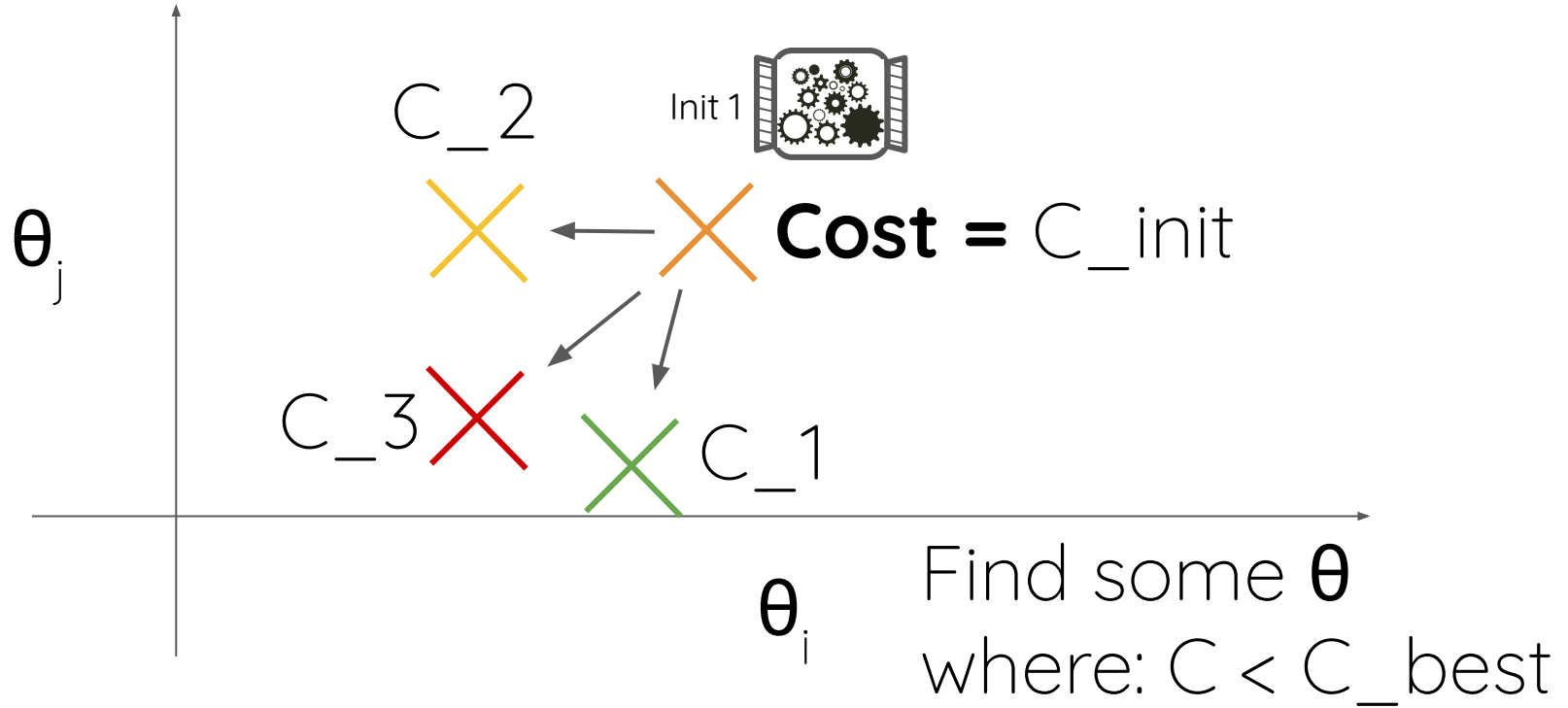
$C(\theta)$



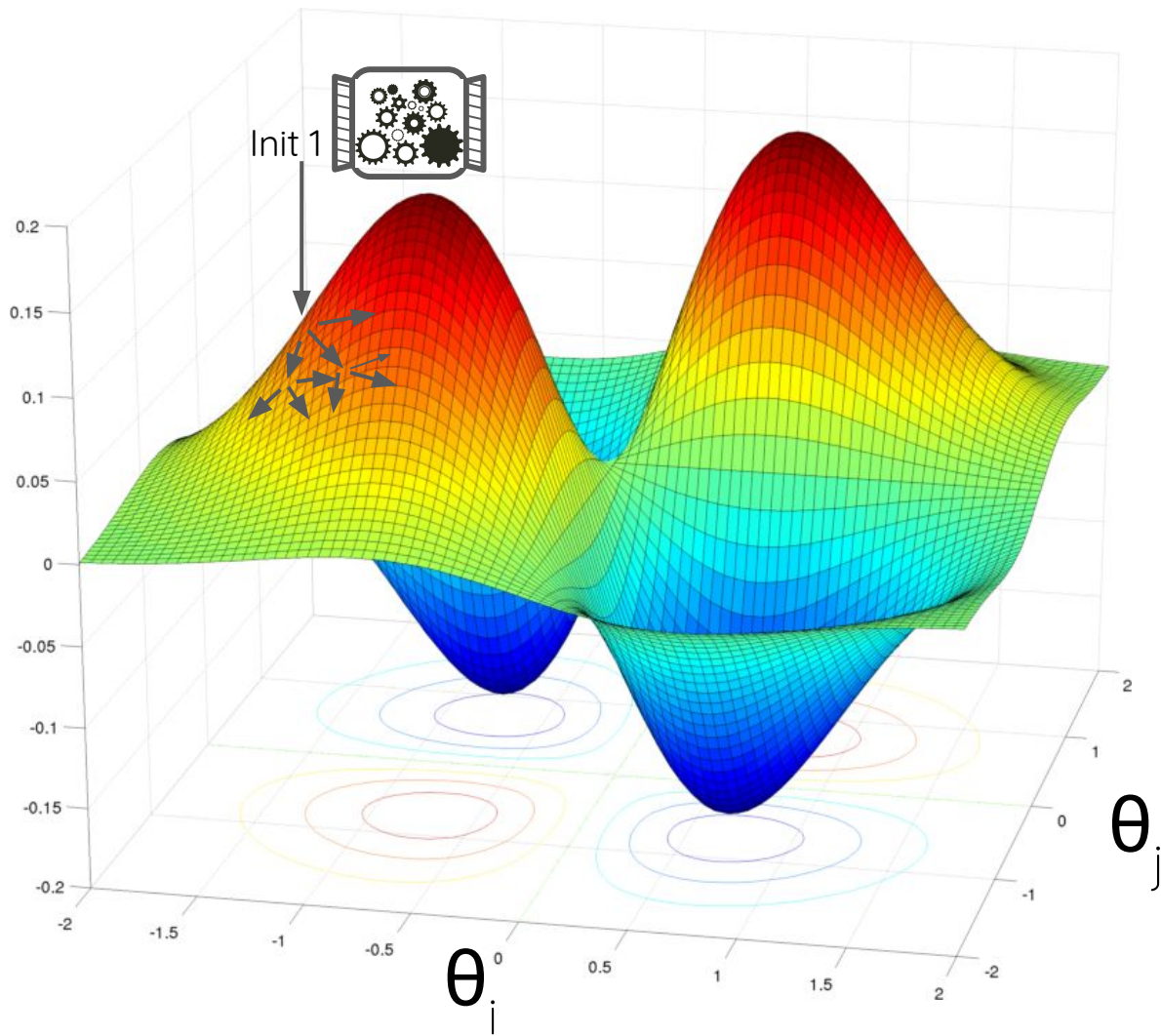
Which way to go?



Next step: Random, visit neighbours?



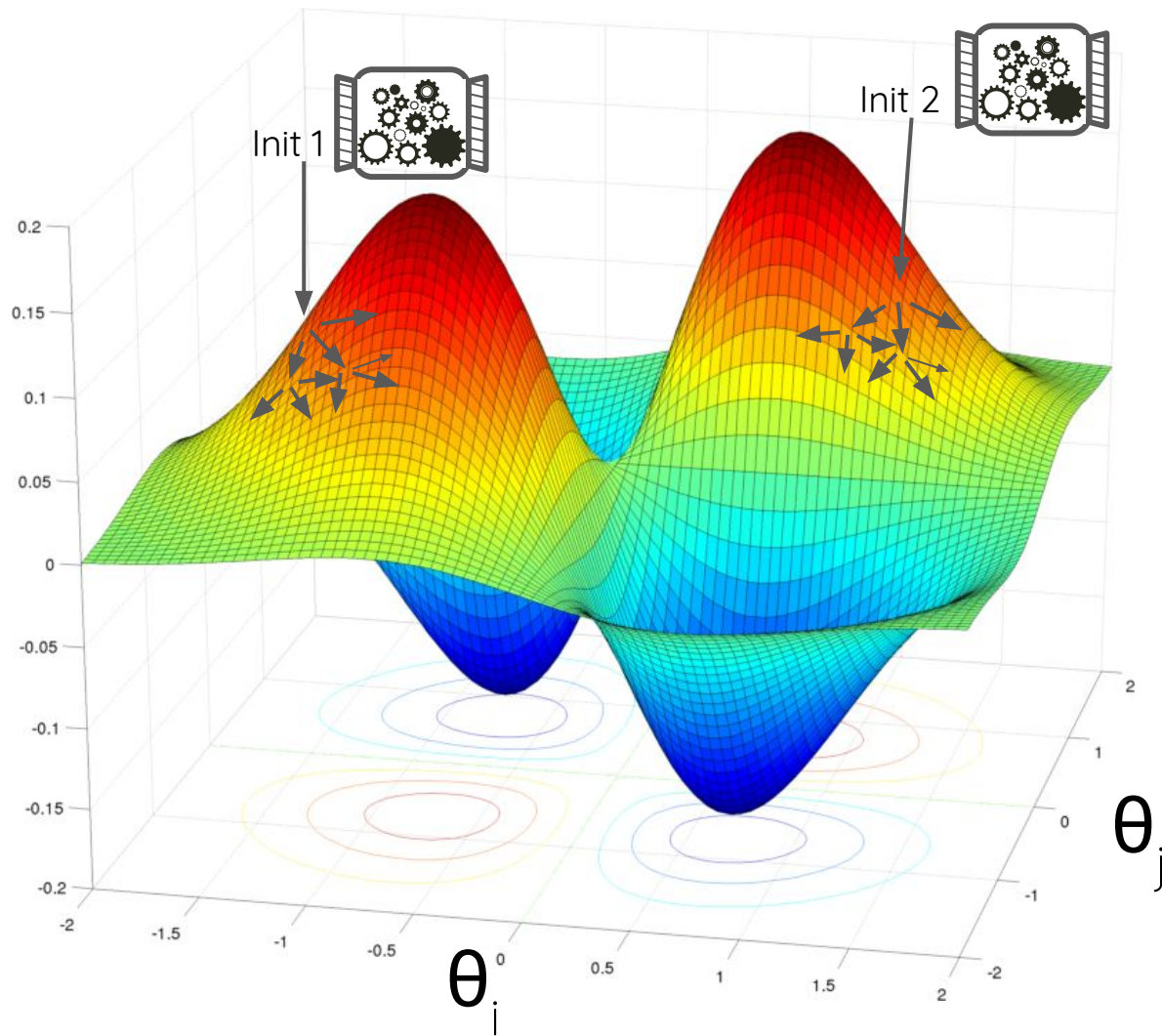
$C(\theta)$



We have no idea about ***f***

Welcome to Monte Carlo

$C(\theta)$

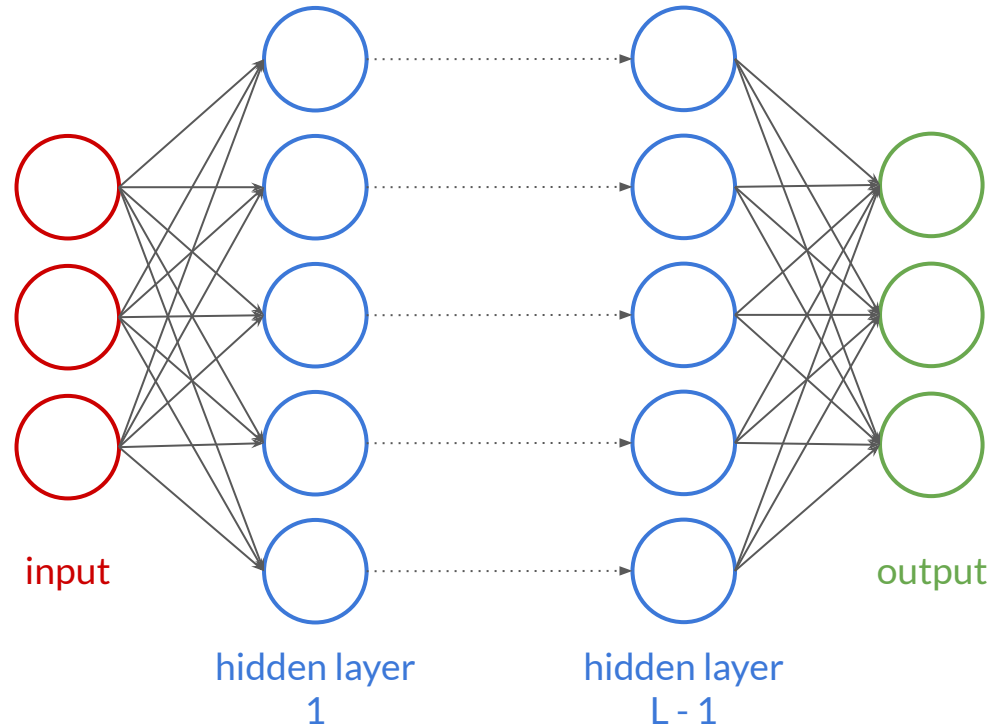


We “have idea” about **f**

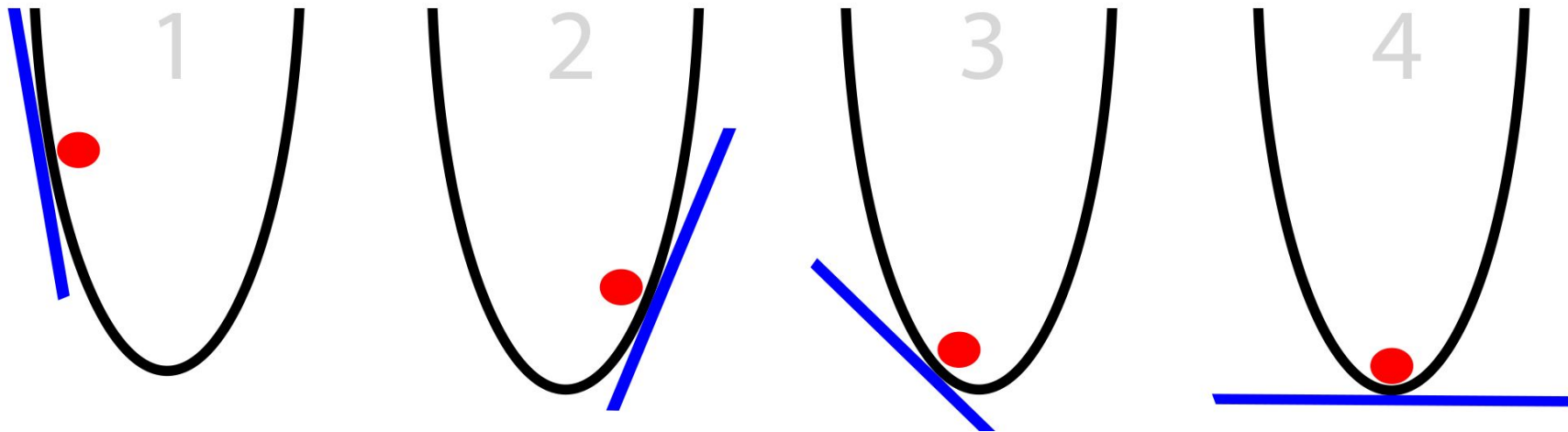
$$\partial C / \partial \theta$$

Gradient descent

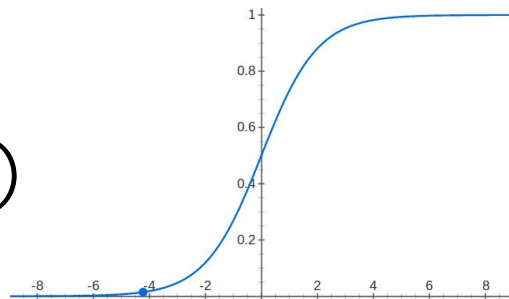
$$a^L \circ a^{L-1} \circ a^{L-2} \circ \dots \circ a^1(x)$$



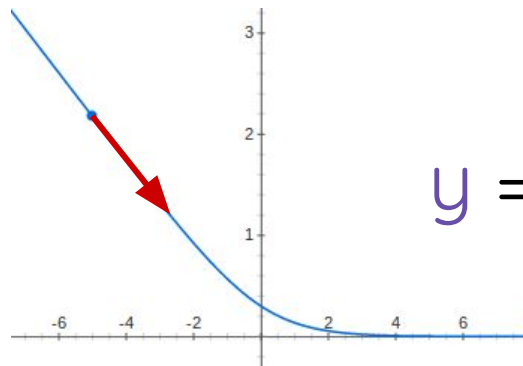
Gradient descent



$$y' = 1 / (1 - \exp(-aL))$$

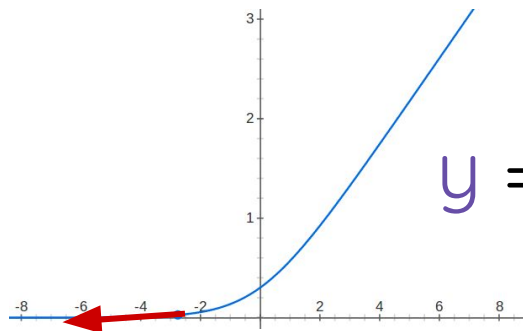


$$-y \log(y')$$



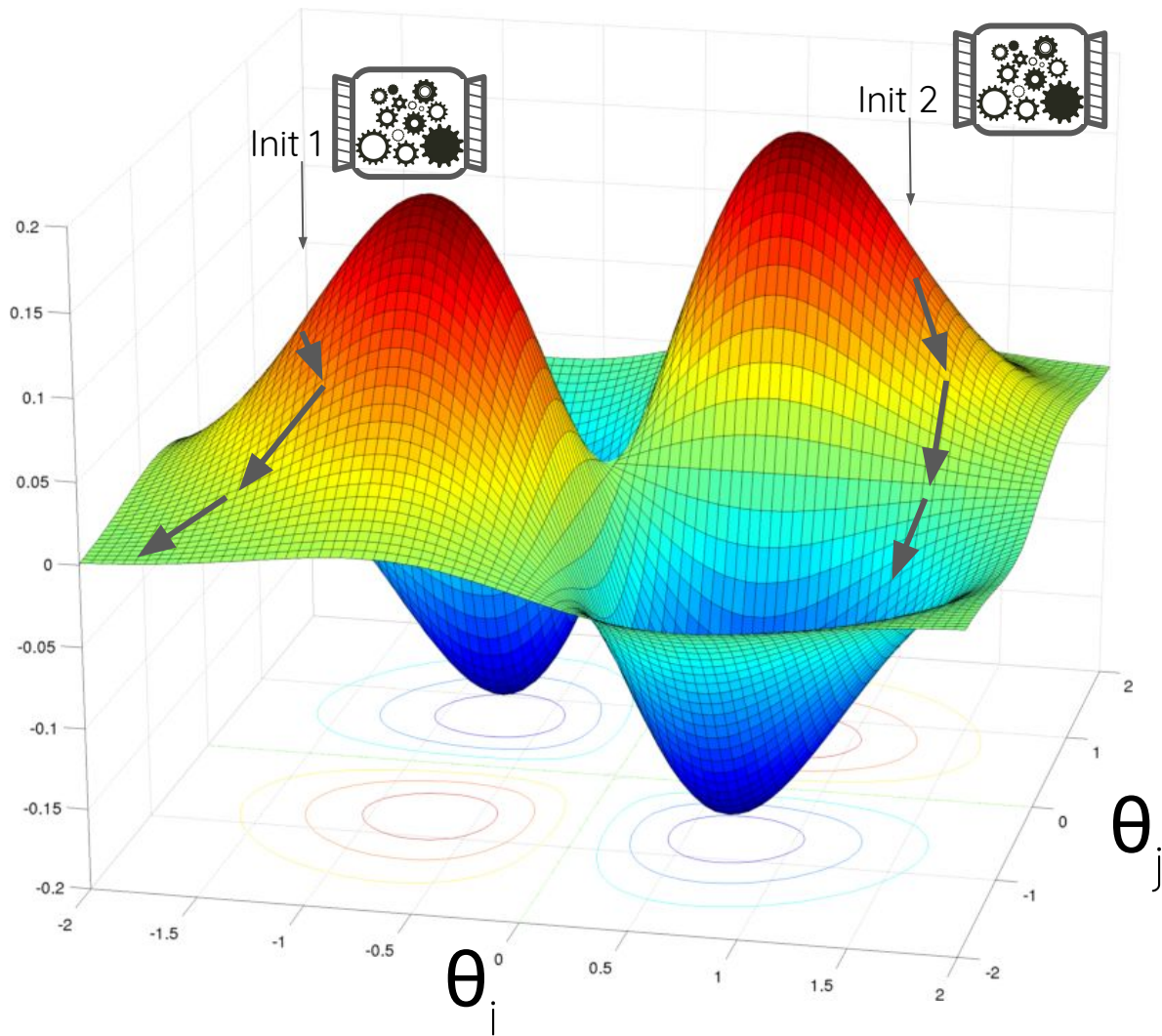
$$y = 1$$

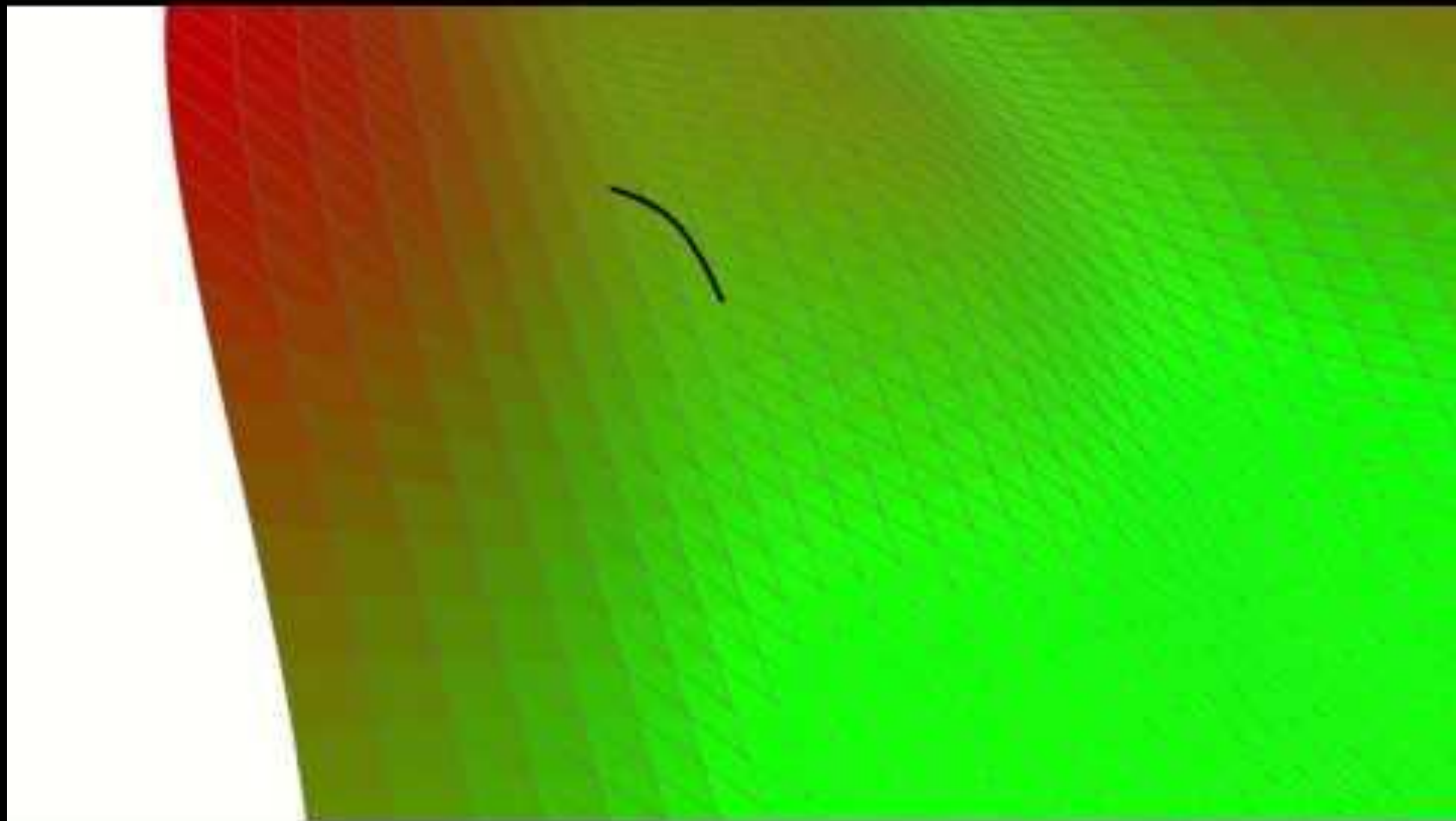
$$-(1-y) \log(1 - y')$$



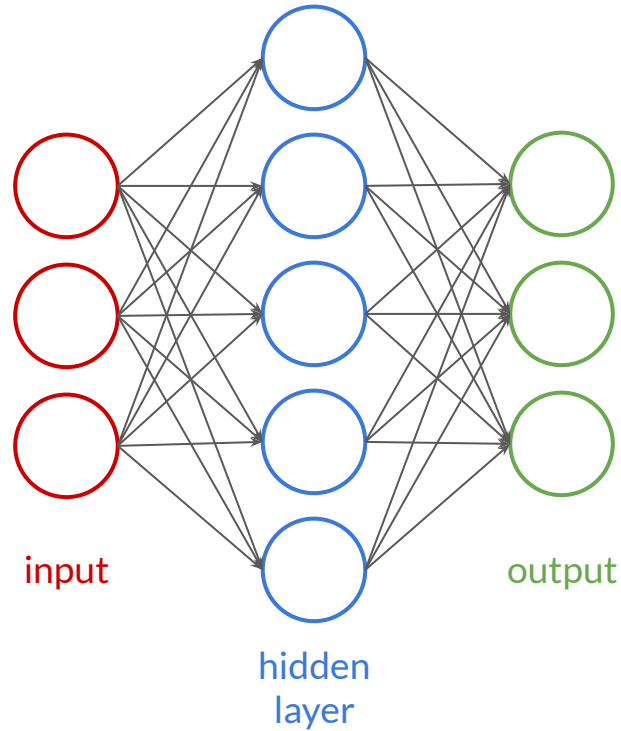
$$y = 0$$

$C(\theta)$

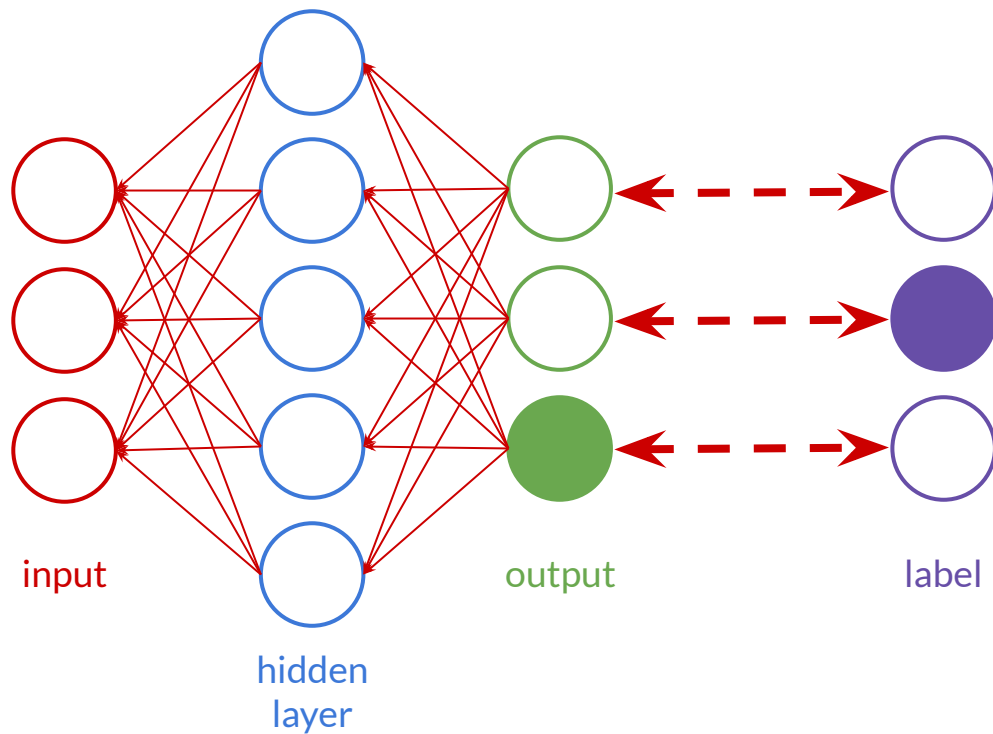




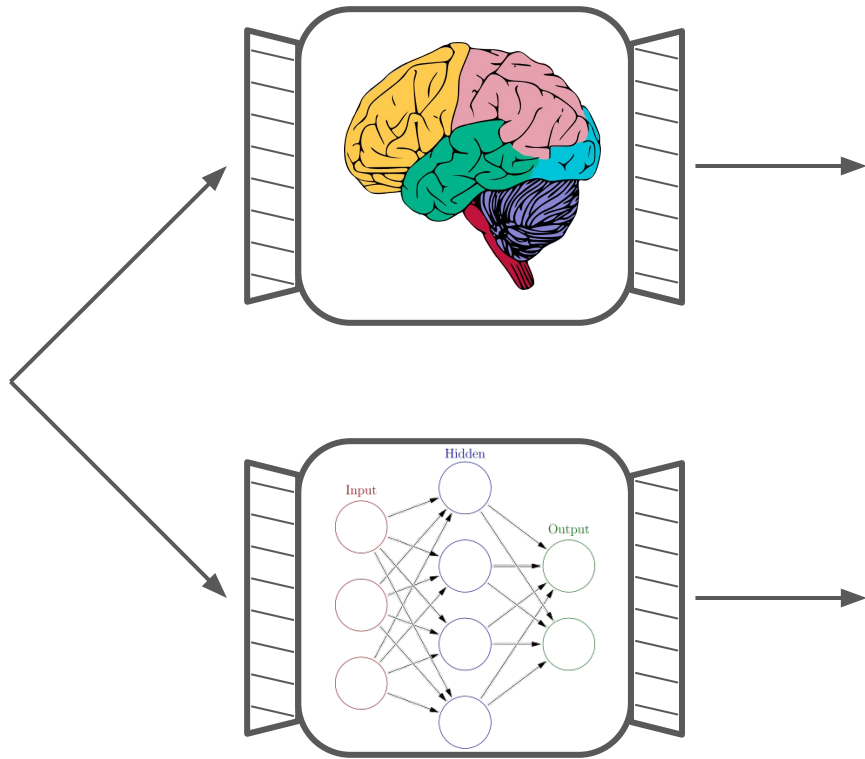
Forward propagation



Backward propagation



X

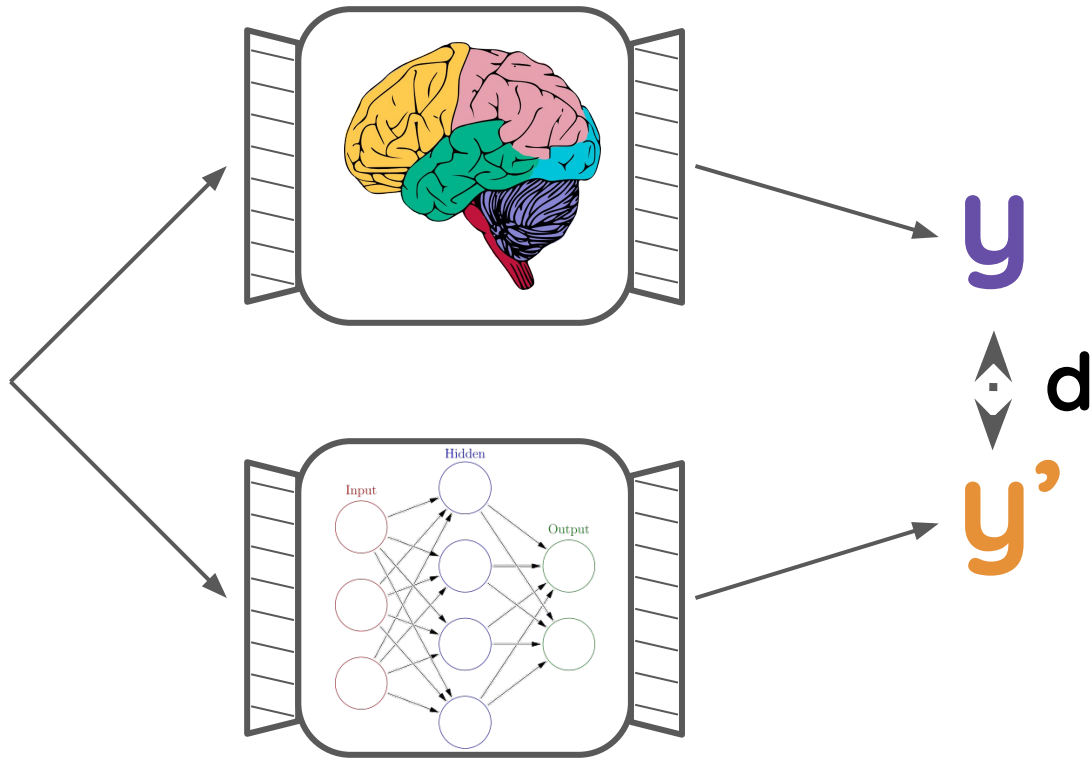


y

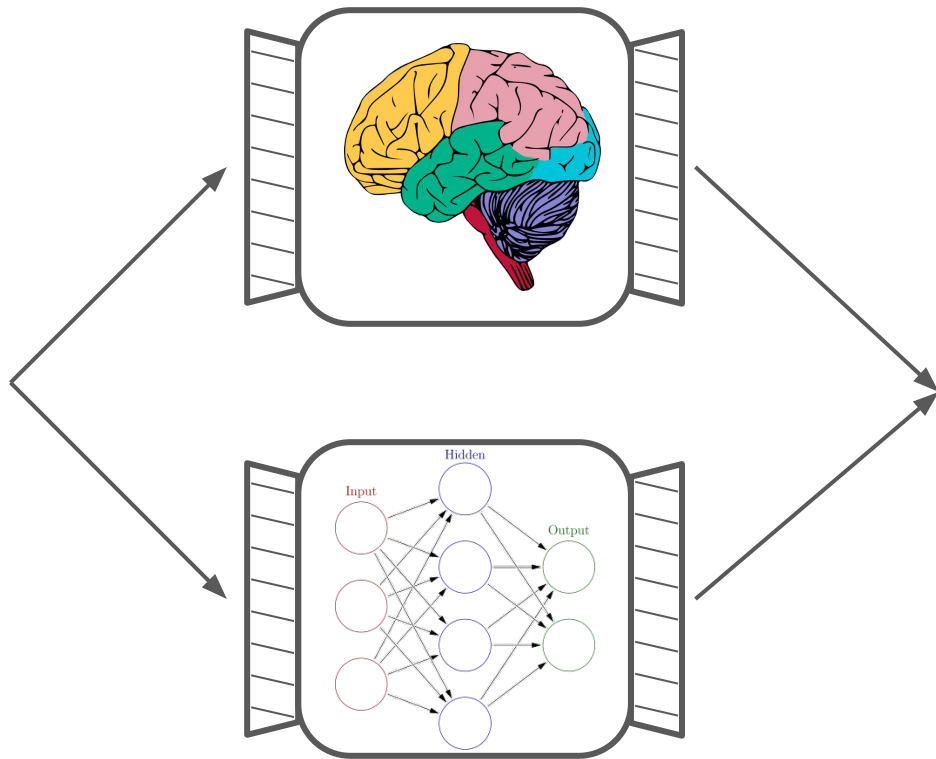
d

y'

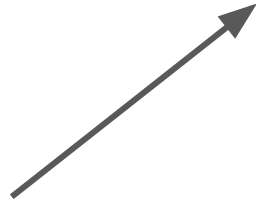
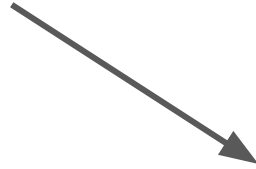
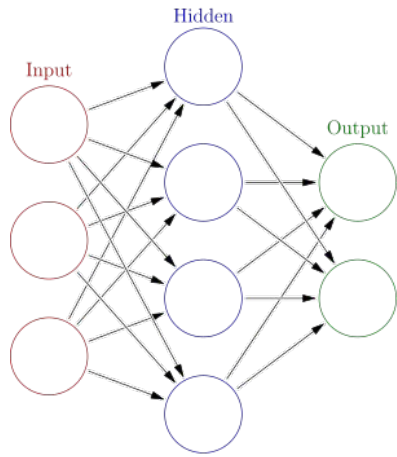
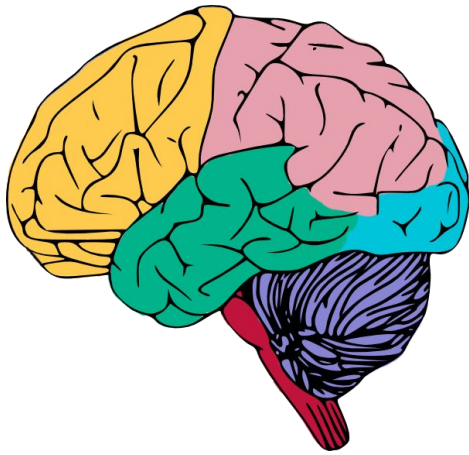
X



x

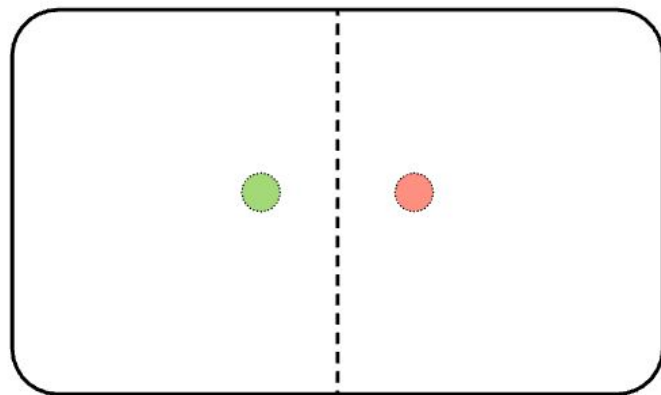


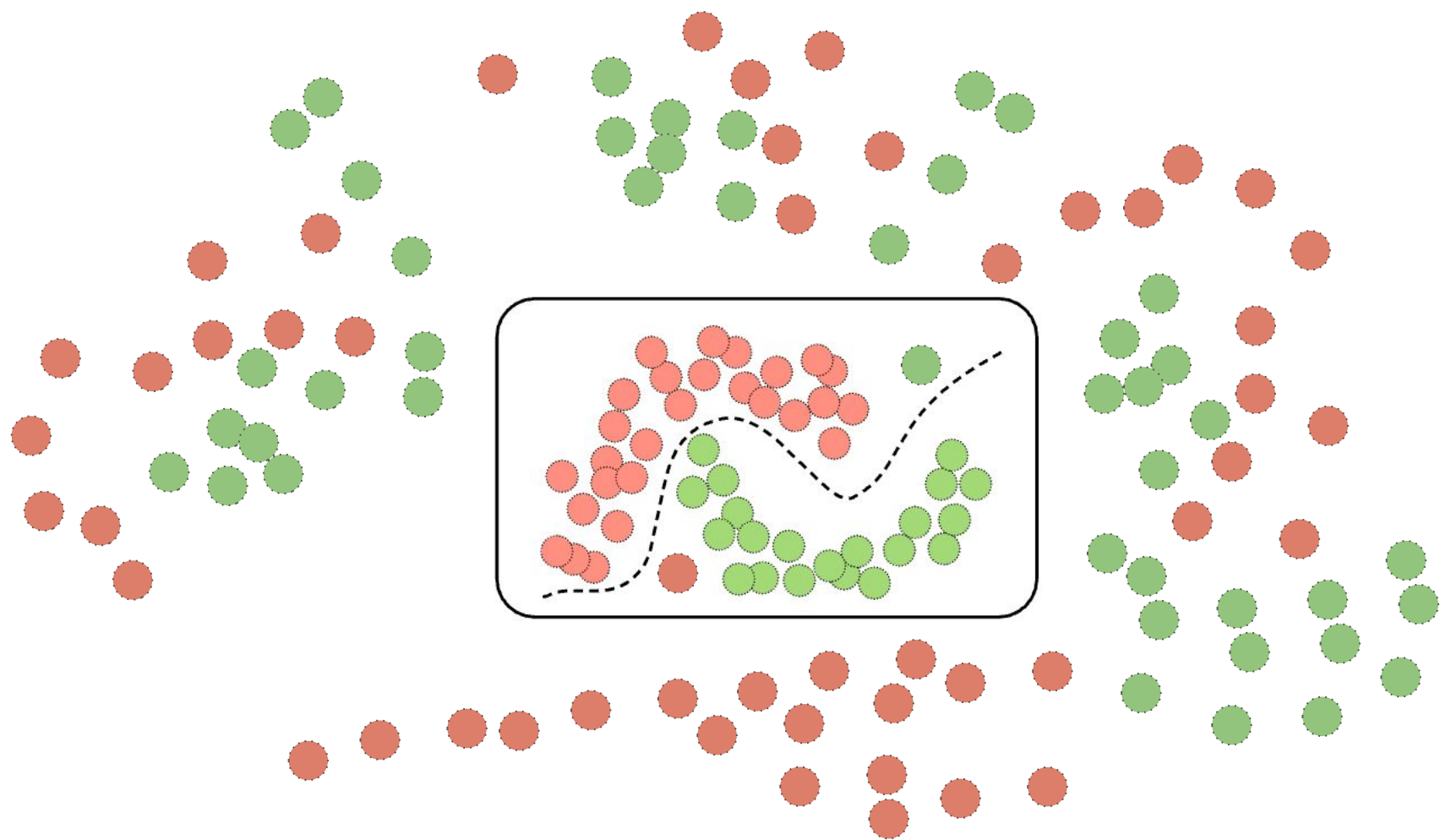
$$y = y'$$

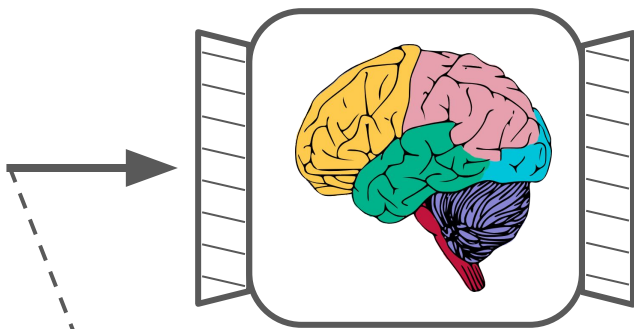
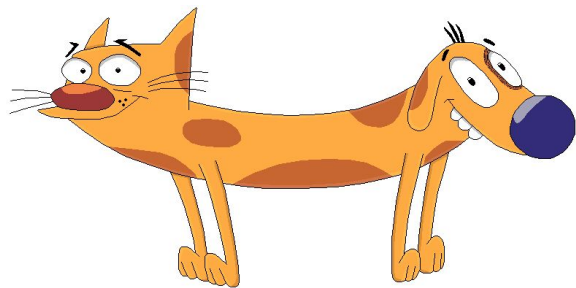


?

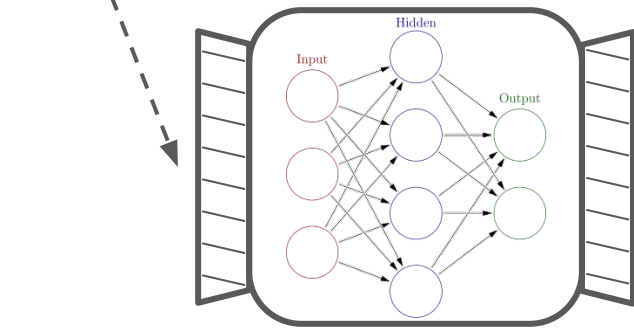






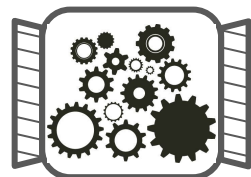
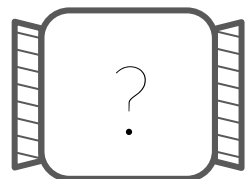


“WTF”



Today's topics

1. Programming or Teaching
2. Parametric functions
3. Distance
4. Optimization (intro)



Köszönöm