SIMPLE NETWORKS

Philipp Krähenbühl

ANNOUNCEMENTS

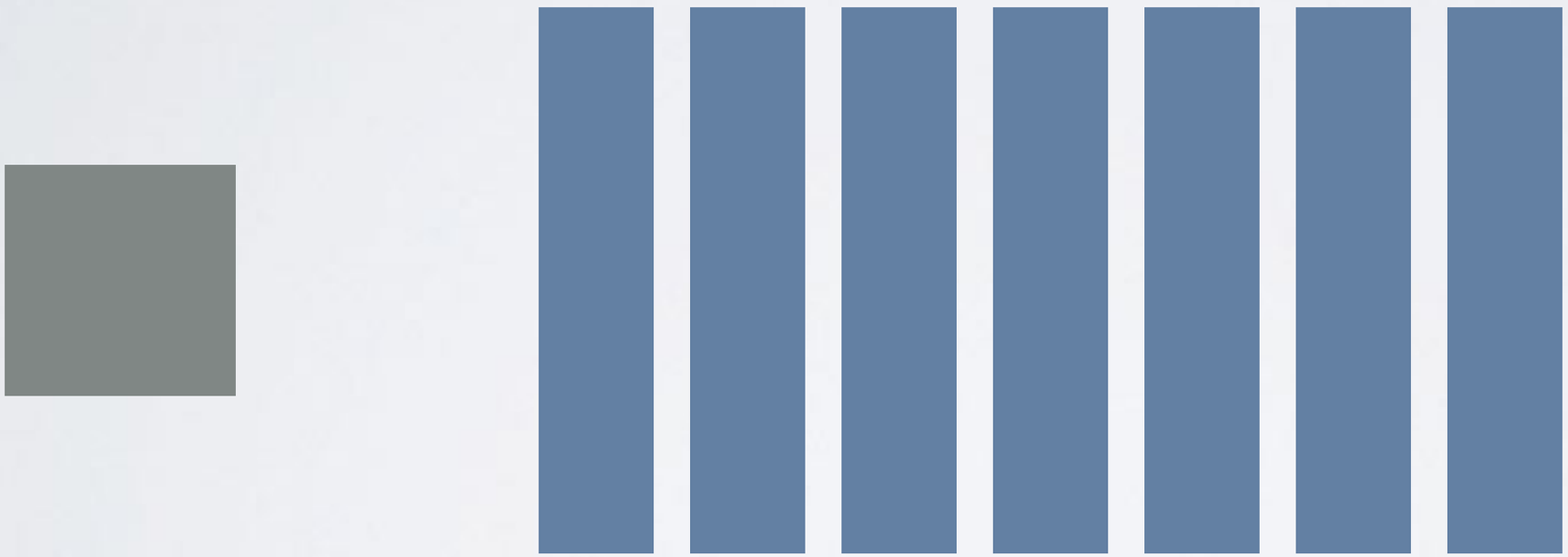
- Start homework 1
  - due Thursday midnight

TODAY

- Linear layer
- ReLU layer
- In class Quiz 1

WHAT IS A DEEP NETWORK?

A "differentiable" function
composed out of multiple layers
of computation



FULLY CONNECTED LAYER

- Affine transformation

$$y = Ax + b$$

- Input
  - n-dimensional vector **x**
- Output
  - m-dimensional vector **y**
- Parameters
  - Linear transformation **A**
  - Bias **b**

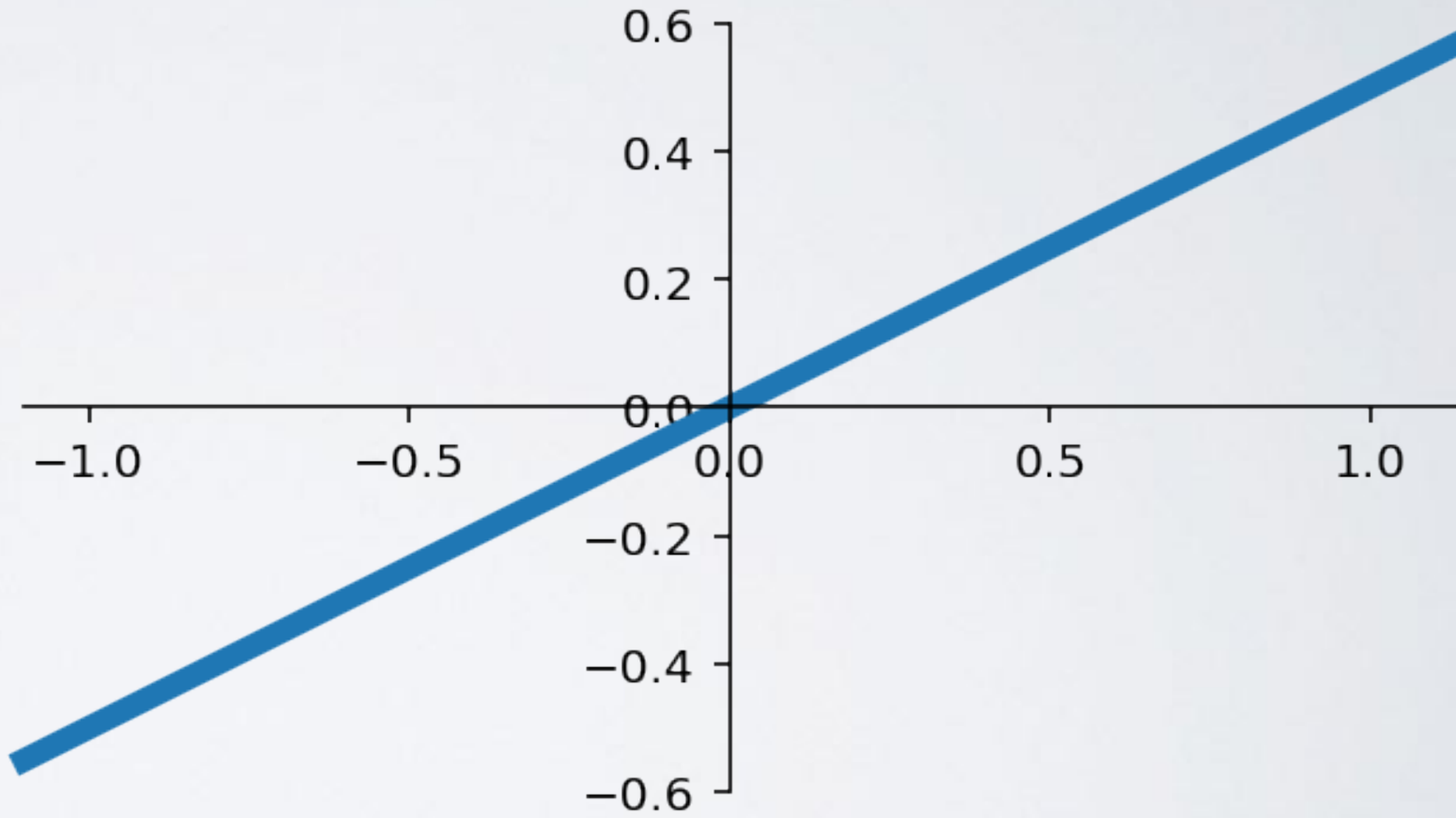
FULLY CONNECTED LAYER

- Affine transformation

$$y = Ax + b$$

- Input
  - n-dimensional vector **x**
- Output
  - m-dimensional vector **y**
- Parameters
  - Linear transformation **A**
  - Bias **b**

What does A control?



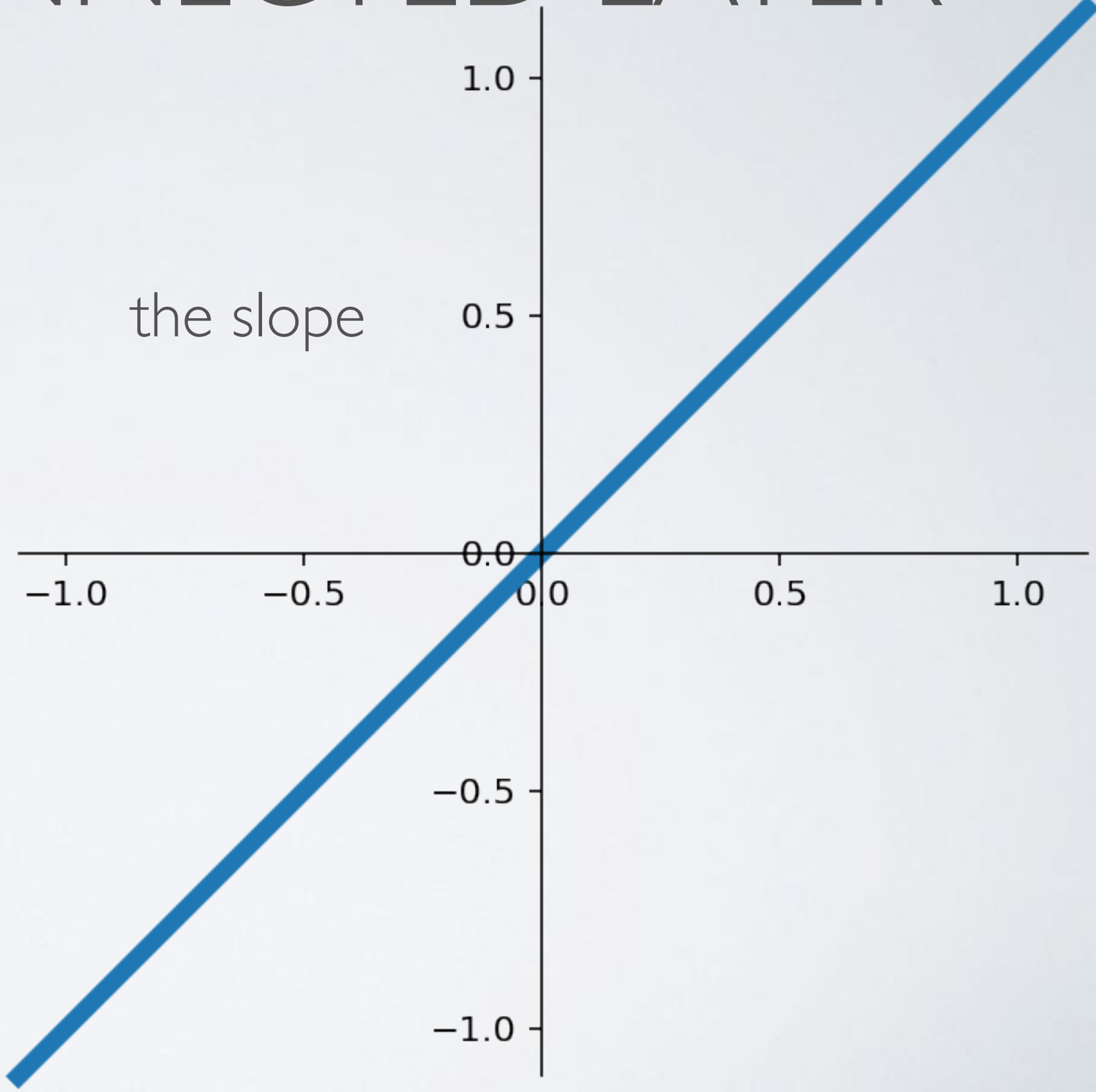
FULLY CONNECTED LAYER

- Affine transformation

$$y = Ax + b$$

- Input
  - n-dimensional vector **x**
- Output
  - m-dimensional vector **y**
- Parameters
  - Linear transformation **A**
  - Bias **b**

the slope



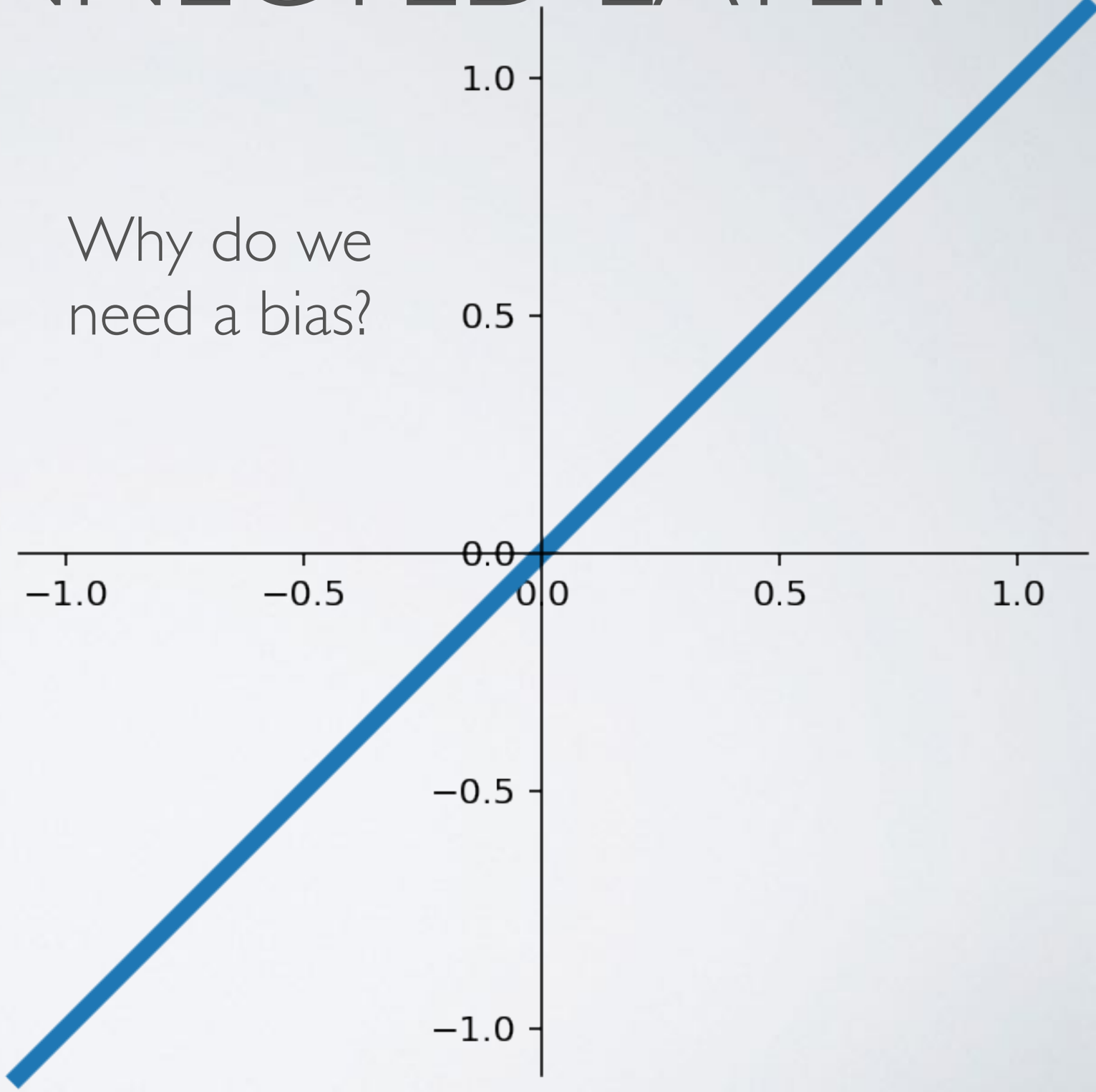
FULLY CONNECTED LAYER

- Affine transformation

$$y = Ax + b$$

- Input
  - n-dimensional vector **x**
- Output
  - m-dimensional vector **y**
- Parameters
  - Linear transformation **A**
  - Bias **b**

Why do we need a bias?



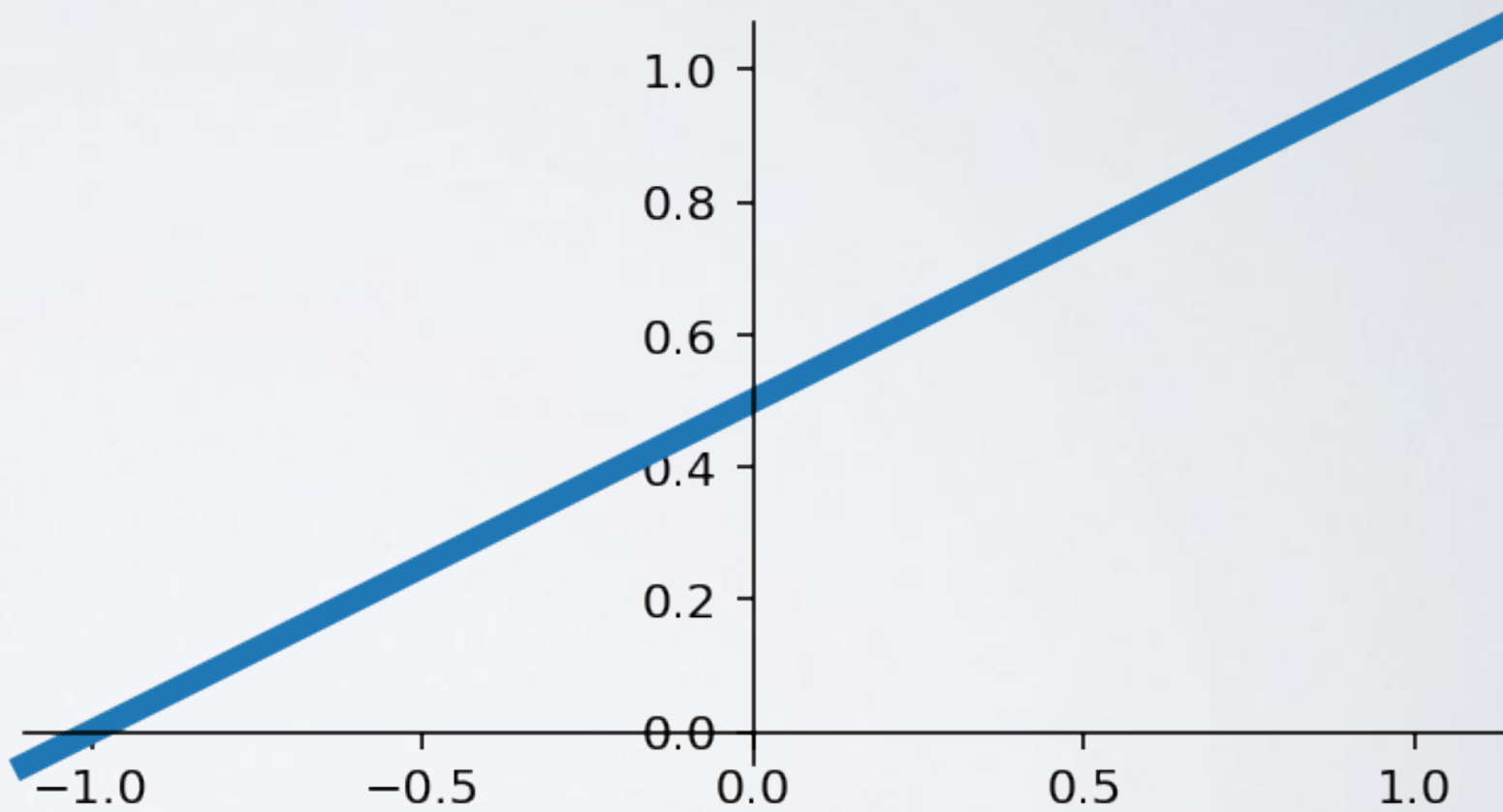
FULLY CONNECTED LAYER

- Affine transformation

$$y = Ax + b$$

- Input
  - n-dimensional vector **x**
- Output
  - m-dimensional vector **y**
- Parameters
  - Linear transformation **A**
  - Bias **b**

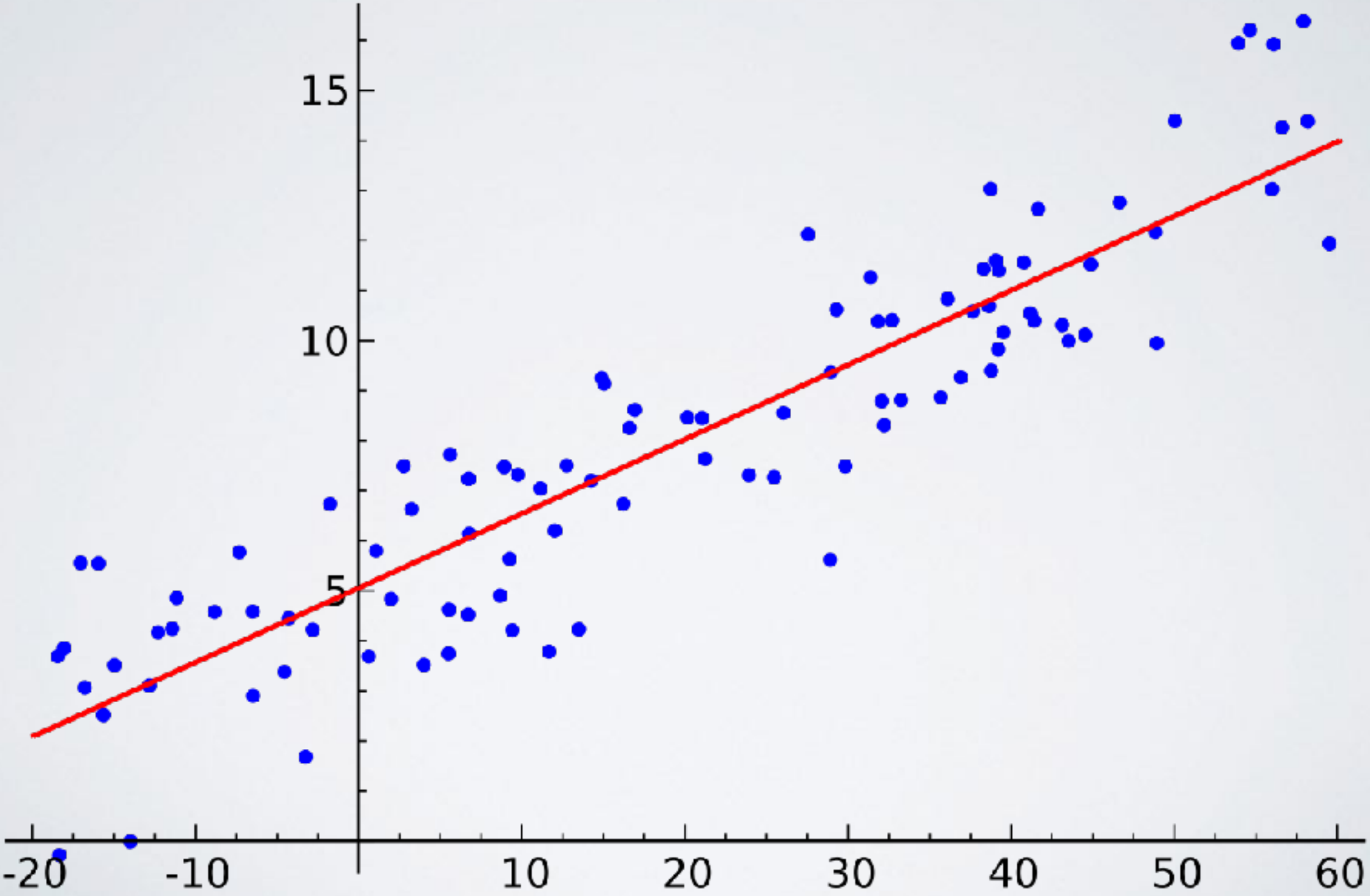
shift the output



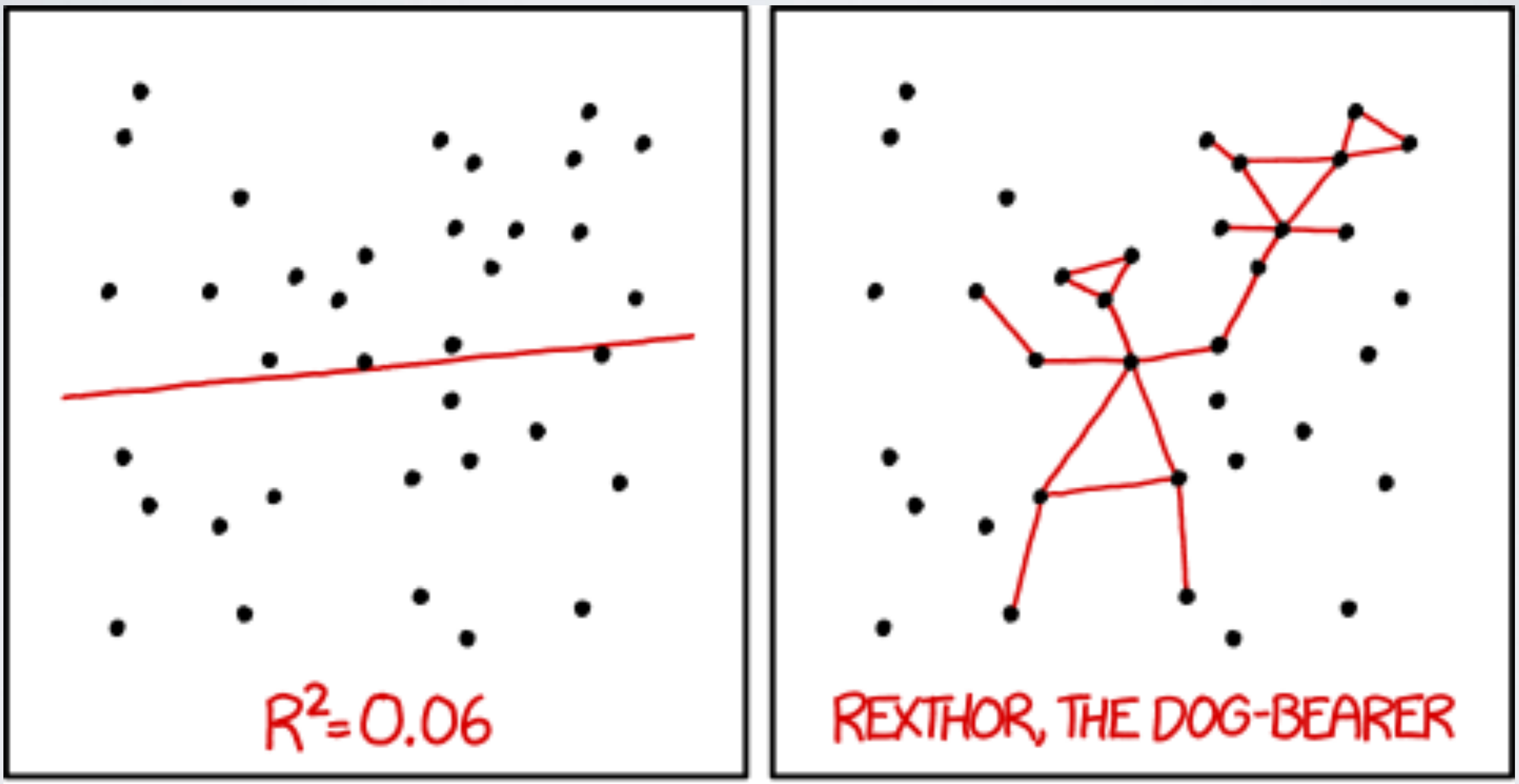
GEOMETRY

- Example
  - Input 2d
  - Output 2d
- Distort points

EXAMPLE: LINEAR REGRESSION



EXAMPLE: LINEAR REGRESSION



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

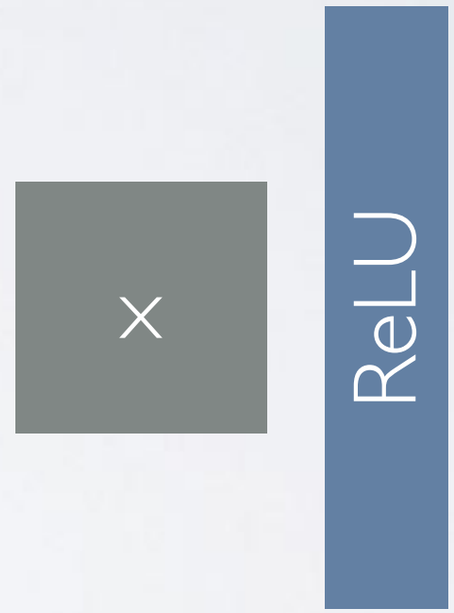
https://xkcd.com/1725/

RELU

- Rectified Linear Unit

$$y = \max(x, 0)$$

- Input
  - tensor **x**
- Output
  - tensor **y** (same dimension as **x**)
- No parameters

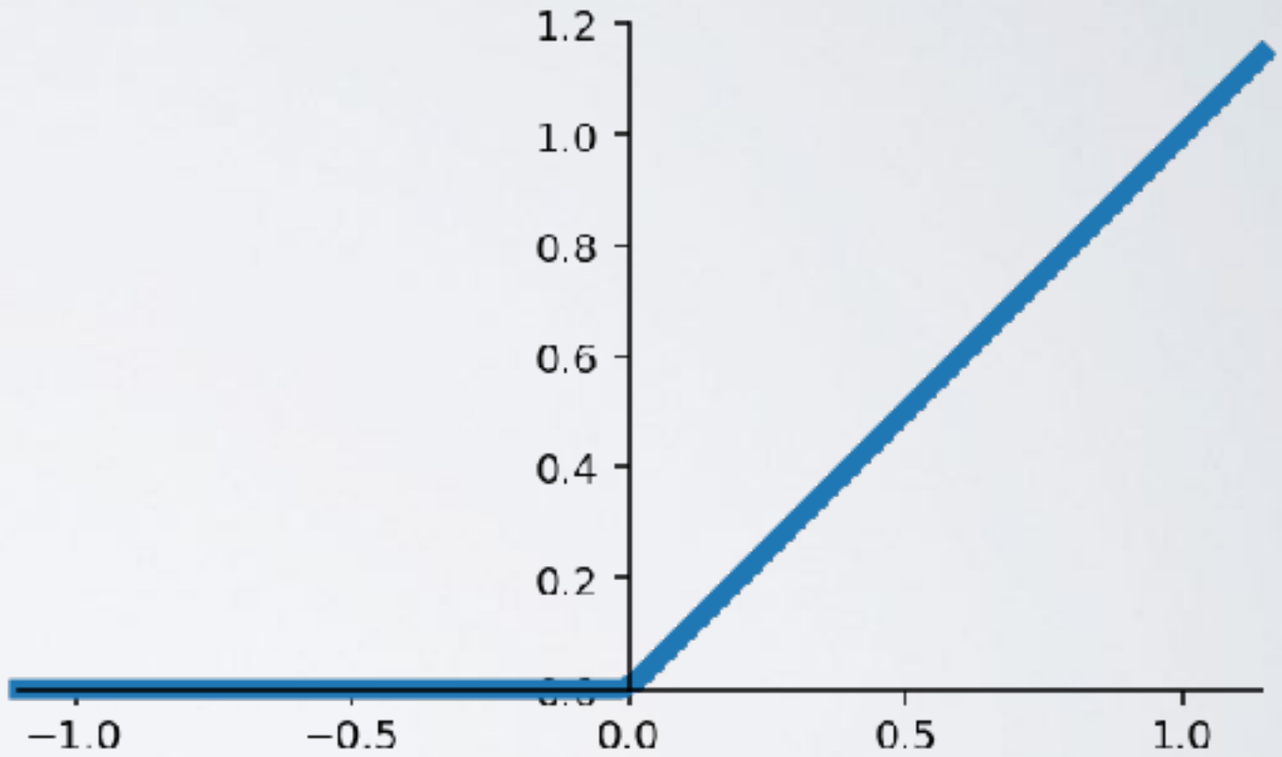


RELU

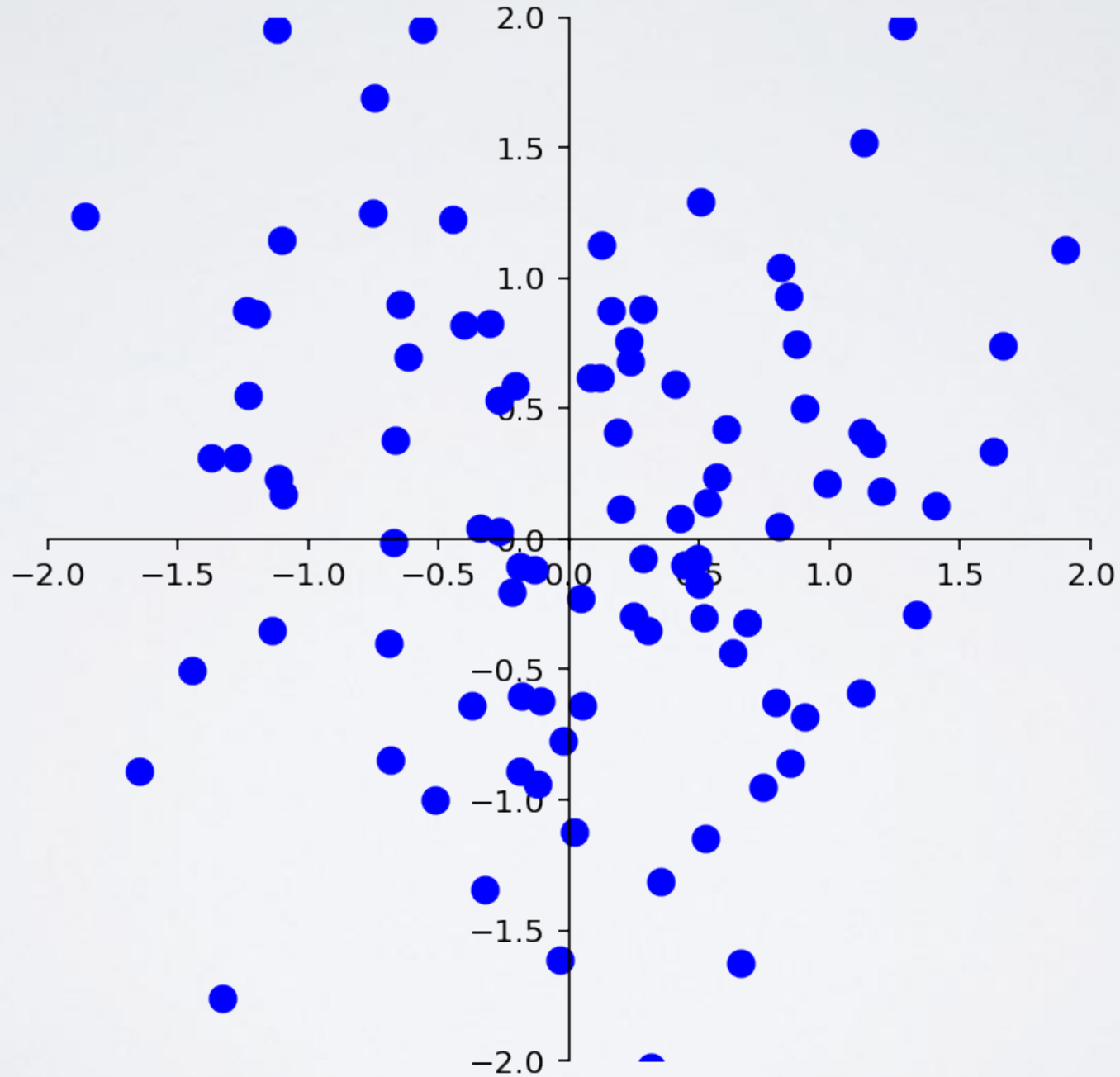
- Rectified Linear Unit

$$y = \max(x, 0)$$

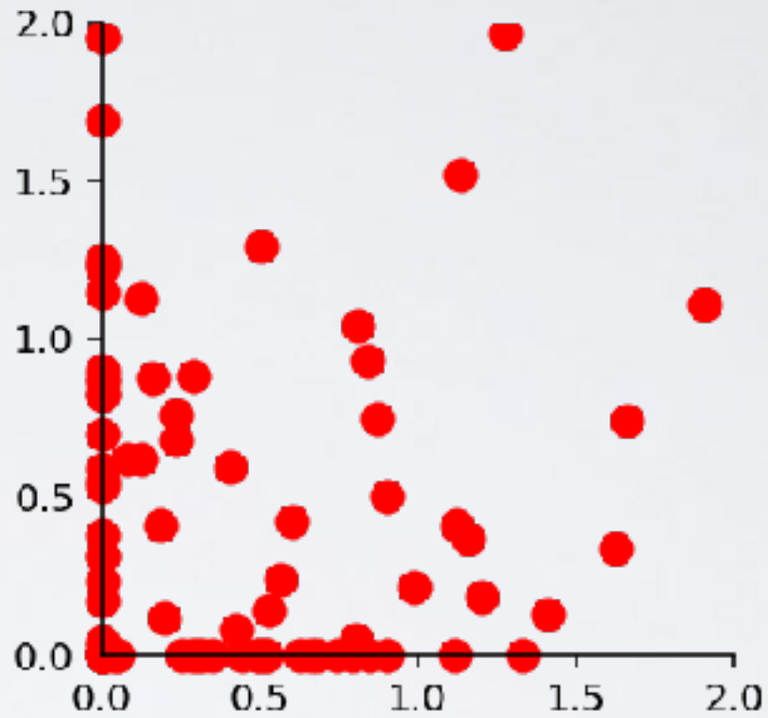
- Input
  - tensor **x**
- Output
  - tensor **y** (same dimension as **x**)
- No parameters



RELU - GEOMETRY

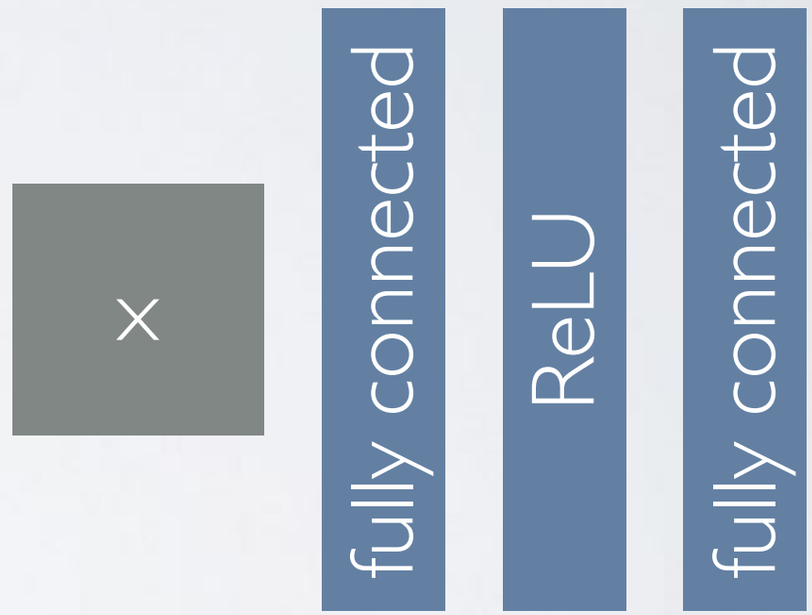


RELU - GEOMETRY



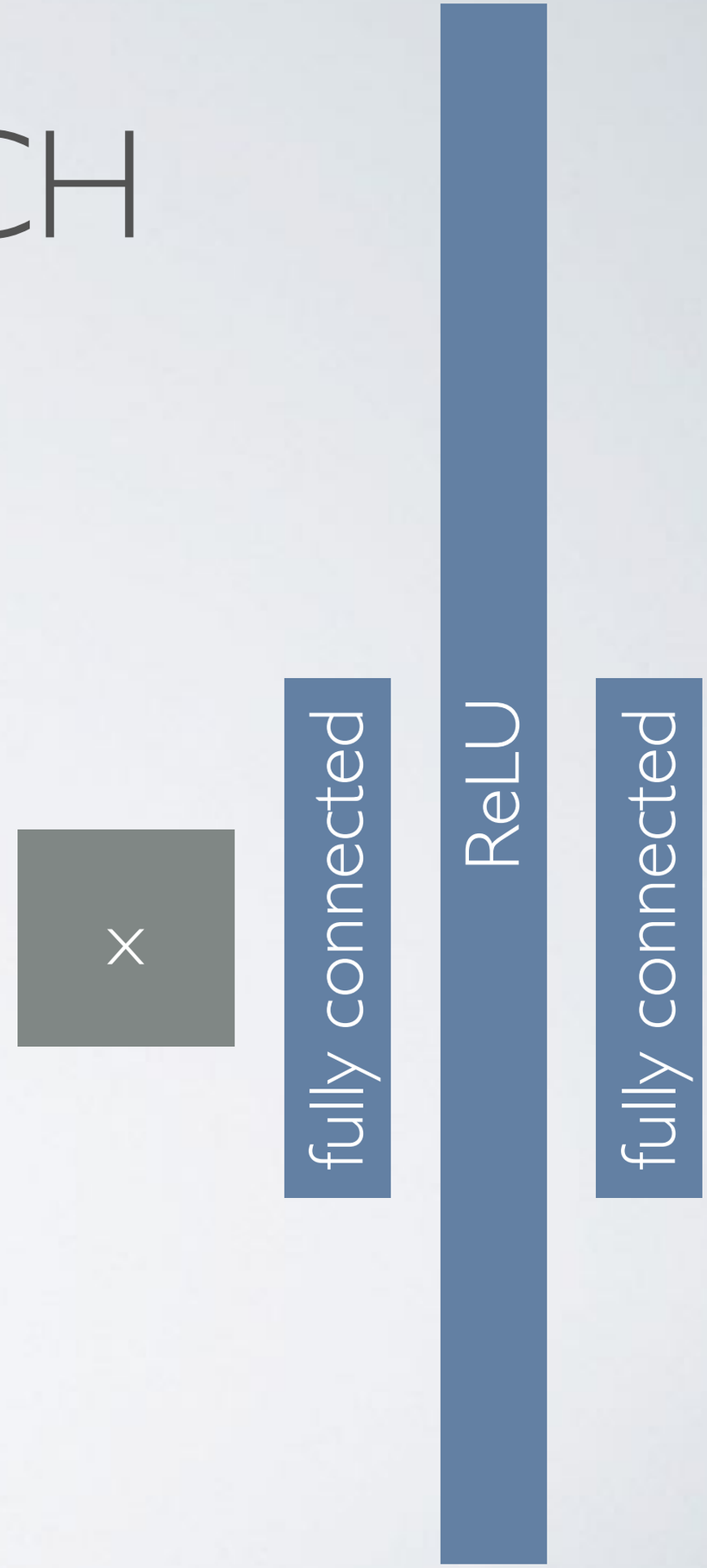
UNIVERSAL APPROXIMATION

- A simply FC-ReLU-FC network can approximate any function on a finite support
  - under mild assumptions
- First proven for Sigmoid (not ReLU) by Cybenko 1989
  - Later extended for ReLU and other non-linearities



THE CATCH

- Universal approximation requires very large fully connected layers
- Very hard to train
  - Computationally
  - Theoretically possible



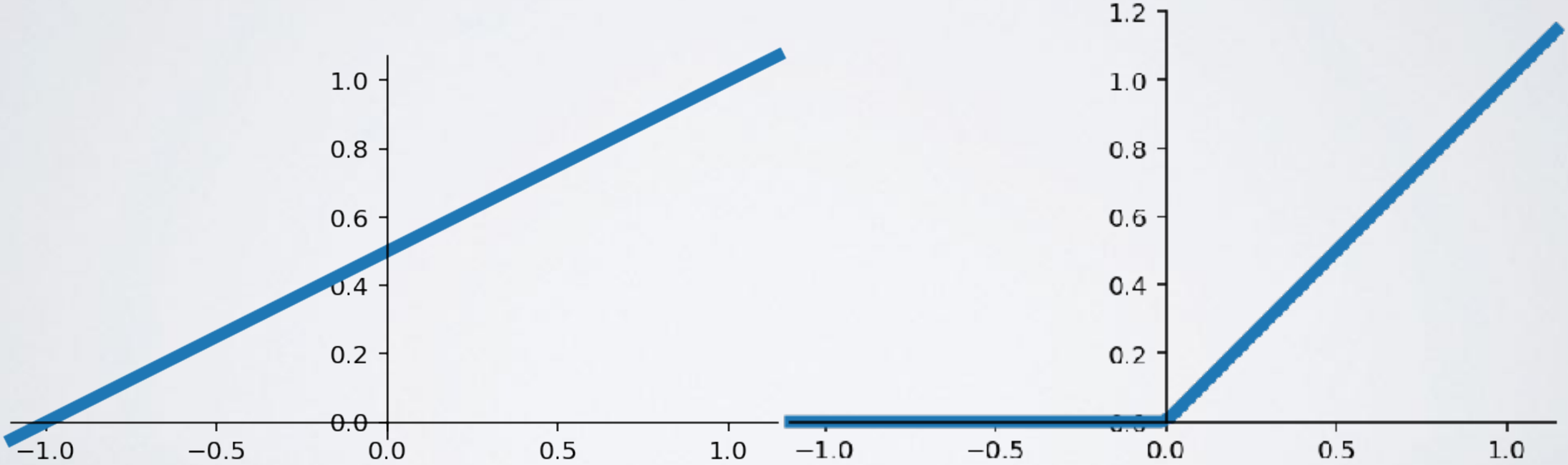
QUIZ: FUNCTION APPROXIMATION

fully connected

$$y = Ax + b$$

ReLU

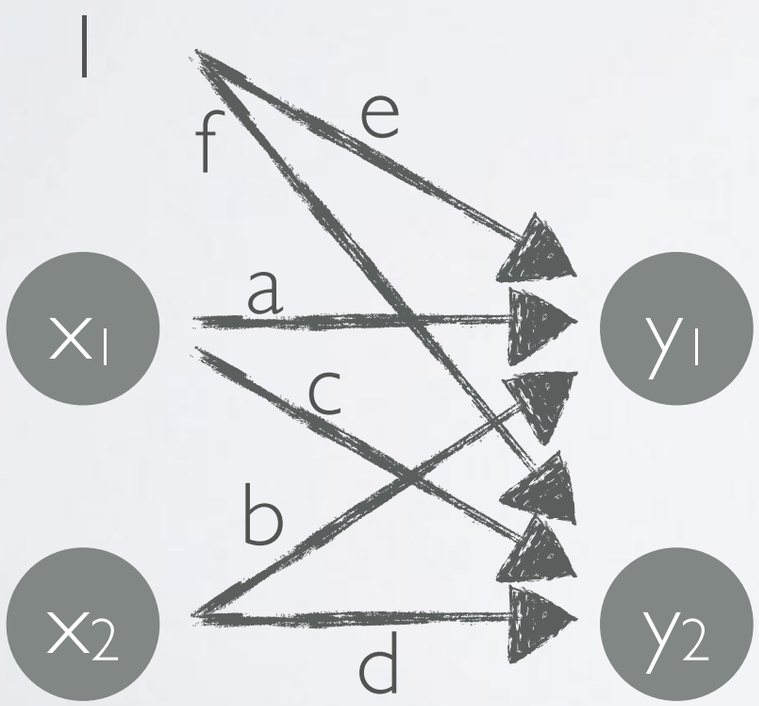
$$y = \max(x, 0)$$



QUIZ: FUNCTION APPROXIMATION

fully connected

$$y = \begin{bmatrix} a & b \\ c & d \end{bmatrix} x + \begin{bmatrix} e \\ f \end{bmatrix}$$



ReLU

$$y = \max(x, 0)$$