Mathematics of Machine Learning / Neural Networks

Corbyn Mellinger

Xu Group Meeting

30 August 2019

What do we mean?

Neural Network

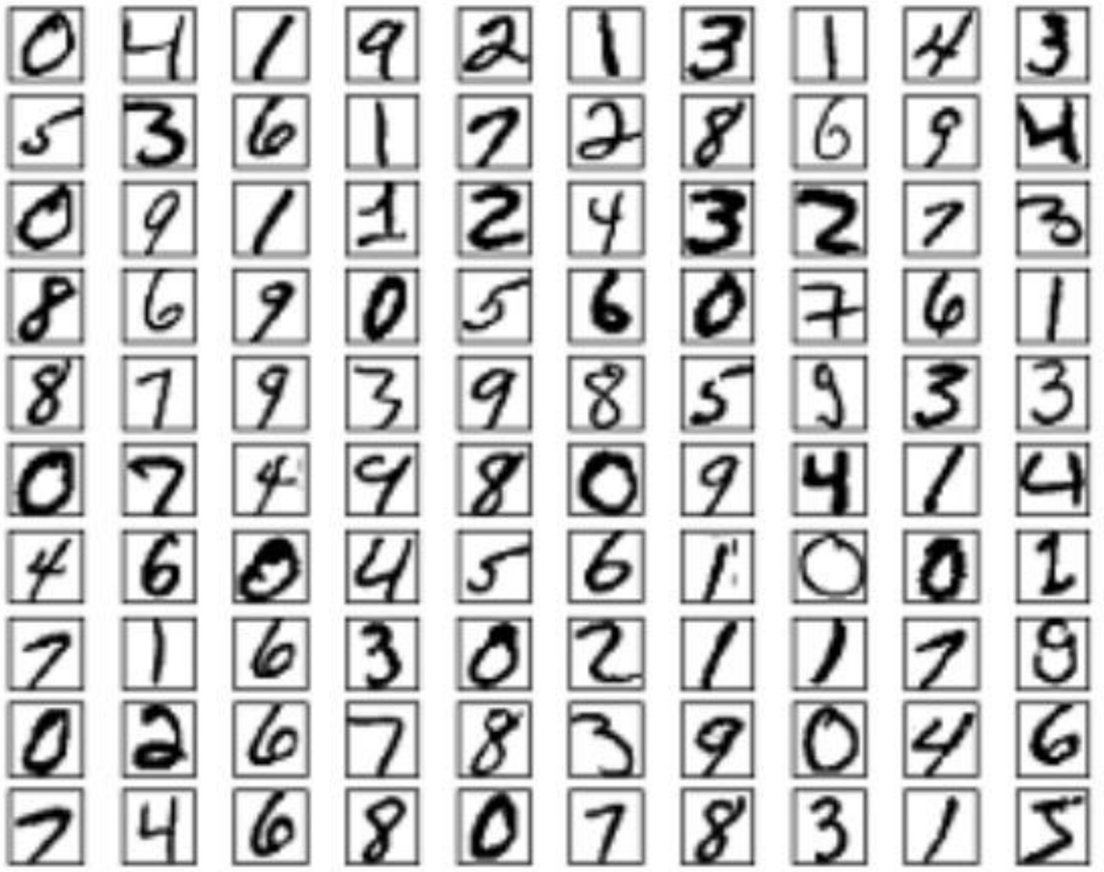
- System of interconnected processes designed (and trained) to perform a task
  - Often, things which currently require human intervention
- Deep Learning: the algorithms used by the neural network

Learning

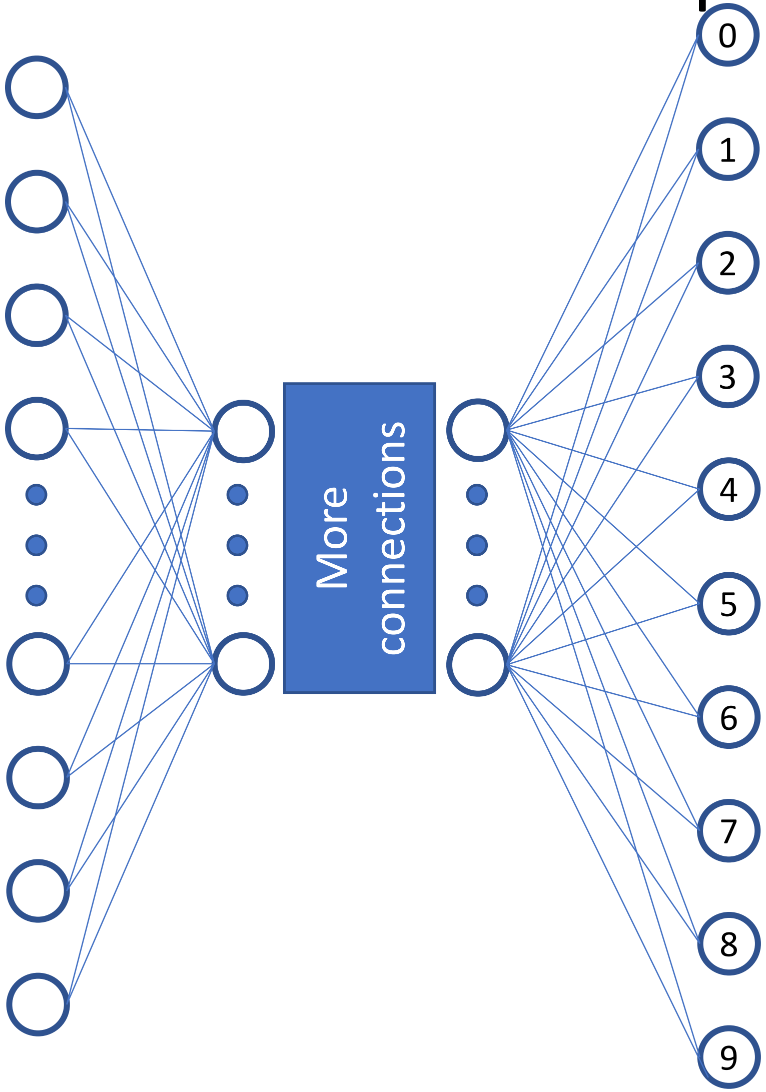
- Similar to human learning: get examples, test knowledge against other examples
- Set up to answer one question **only**
  - Will give answer, but not always necessarily sensible

Common Example: Identify numbers

- Need to identify numbers which deviate from a “perfect” computer font style
- Network: input pixel values, (go through some processing), output the number it is (with some indicator of certainty)



Common Example: Identify numbers

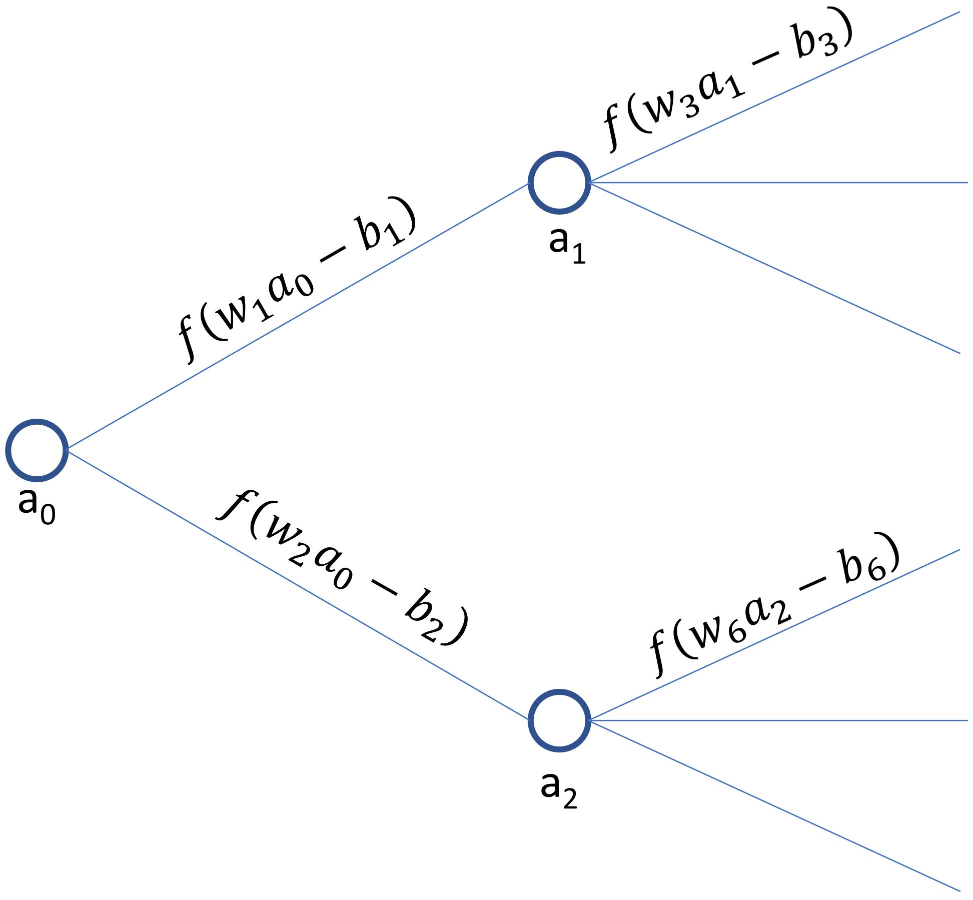


N x N pixel inputs

Output of what it thinks # is

Connections chosen based on algorithm, process you envision network taking

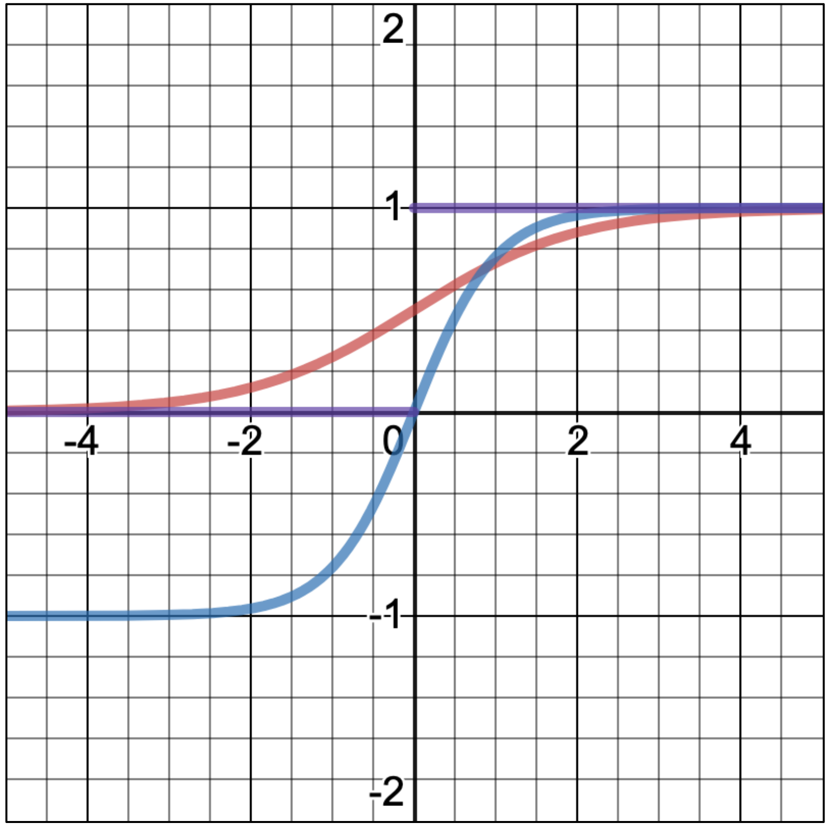
What’s in a connection?



Goal: optimize these weight, bias parameters to give most accurate output

f: distribution function which takes input, weights it to each output, and “biases” by some amount

e.g. sigmoid, tanh, step functions



“Training” Schedule

1. Input numbers from database
2. Allow system to output
3. Tell system the correct answer, quantify how wrong it was
4. Repeat for many sets…

Many choices for quantification: consider one we all already know, difference of squares

$$C = \sum_{i\ outputs} (a_L - y)^2,$$

y is desired output

Minimize this function of parameter space

→ Do calculus to minimize C

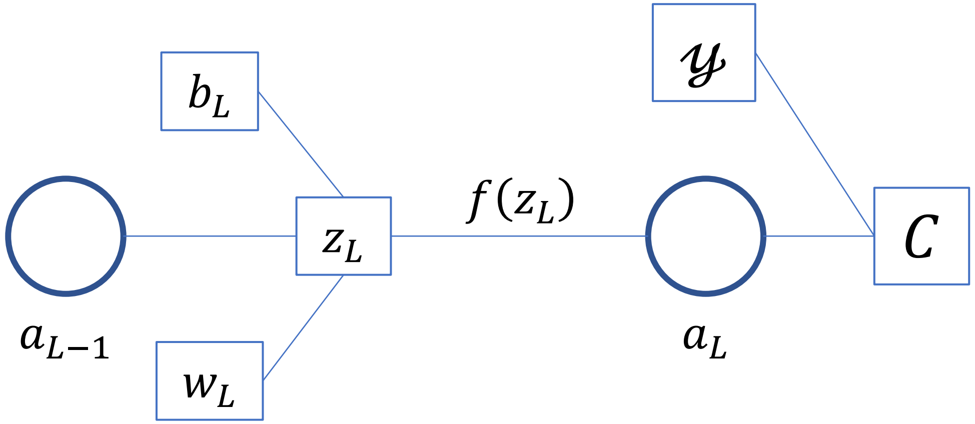
Simple example: one connection

$$\frac{\partial C}{\partial w_L} = \frac{\partial C}{\partial a_L}\frac{\partial a_L}{\partial z_L}\frac{\partial z_L}{\partial w_L}$$

$$C = (a_L - y)^2 \Rightarrow \frac{\partial C}{\partial a_L} = 2(a_L - y)$$

$$a_L = f(z_L) \Rightarrow \frac{\partial a_L}{\partial z_L} = f'(z_L)$$

$$z_L = w_L a_{L-1} + b_L \Rightarrow \frac{\partial z_L}{\partial w_L} = a_{L-1}$$



Find how C depends on *all* parameters, then find “gradient of steepest descent” to get to a minimum in the cost function

Other Resources

- Good (free) online textbook: http://neuralnetworksanddeeplearning.com/index.html

- Series of YouTube videos: https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi