Introduction to Deep Learning

Ups and Downs of Deep Learning

- 1958, perceptron (linear model)
- 1969: Perceptron has limitation
- 1980: Multi-layer perceptron (X)
  - Do not have significant difference from DNN today
- 1986: Backpropagation
  - Usually more than 3 hidden layers is not helpful
- 1989: 1 hidden layer is 'good enough', why deep learning?
- 2006: RBM initialization (breakthrough)
- 2009: GPU
- 2011: Start to be popular in speech recognition
- 2012: win ILSVRC image competition

Three Steps for Deep Learning

- Define a set of function (Neural Network)
- Evaluate the performance of function
- Pick up the best function

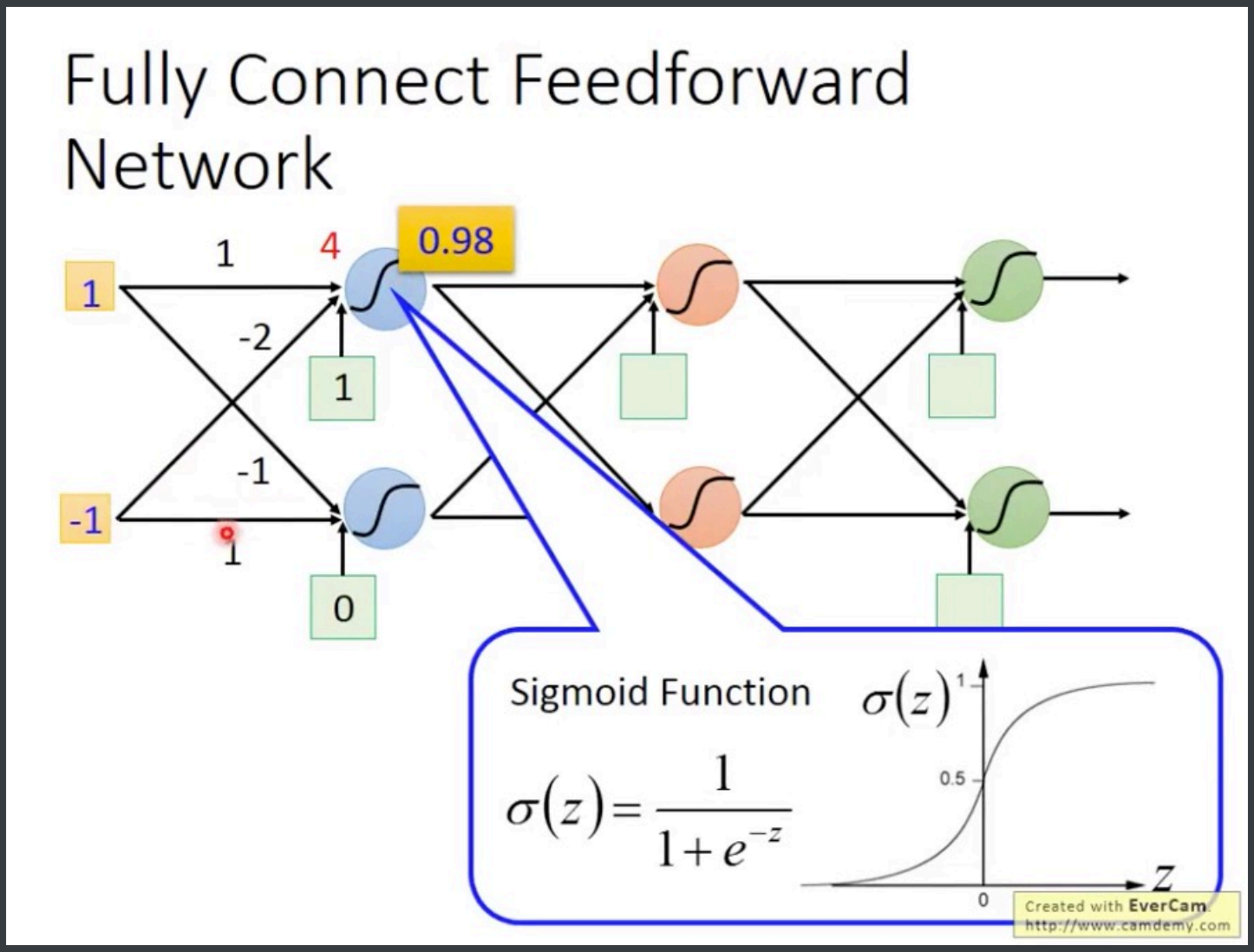
Neural Network

Concatnate different logistic regressions, each logistic regression has its own weight and bias.

Different connection leads to dfferent network structures:

Network parameter theta: all the weights and biases in the 'neurons'

How to connect them?



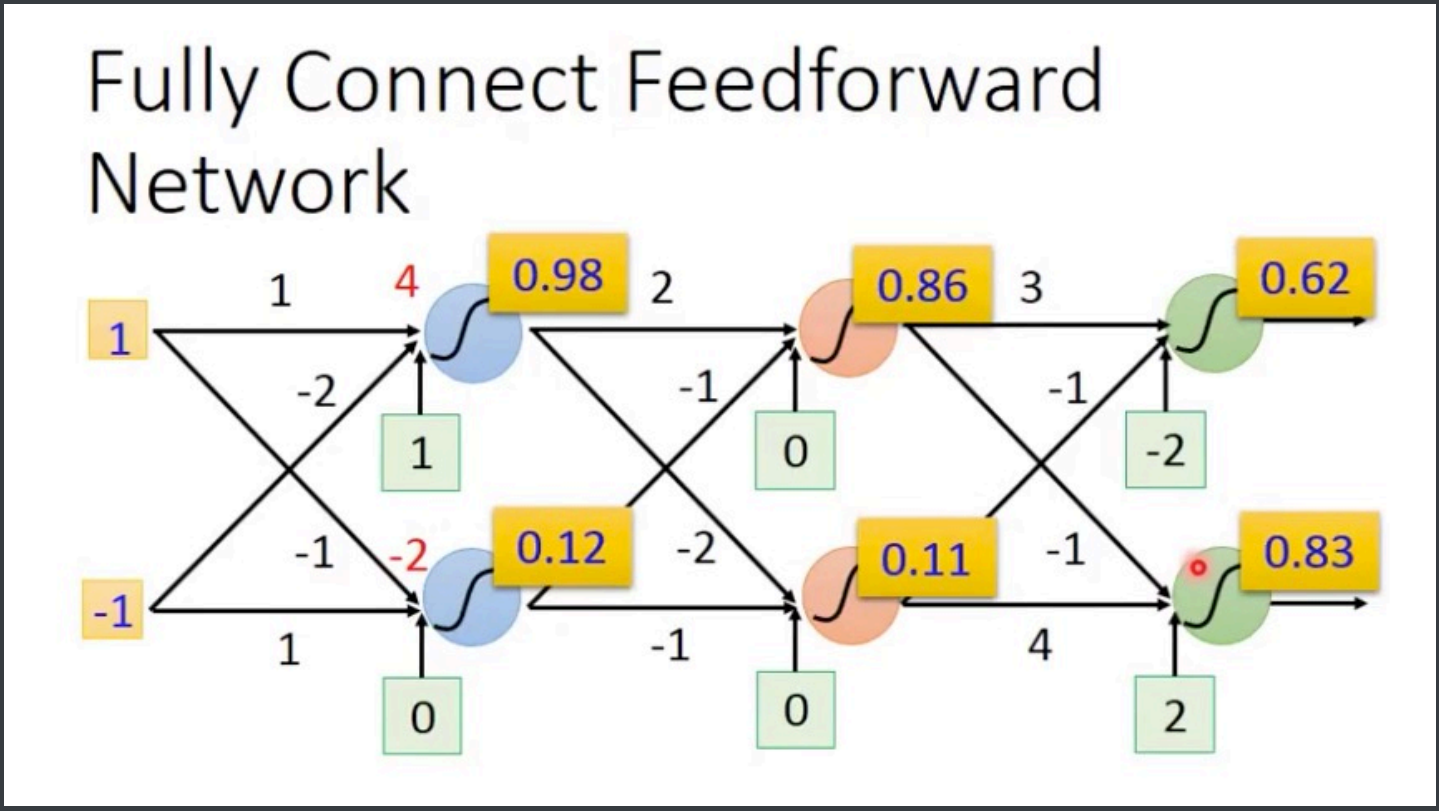
weight and bias are from training data.

for example, blue top neuron, weight (1, -2), bias(1); blue bottom neuron, weight (-1, 1), bias(0)

top blue output: 1x1 + (-1) x (-2) + 1 (bias) = 4 >>> Pass the sigmoid function >>> 0.98

bottom blue output: 1x(-1) + (-1)x1 + 0 = -2 >>> Pass the sigmoid function >>> 0.12

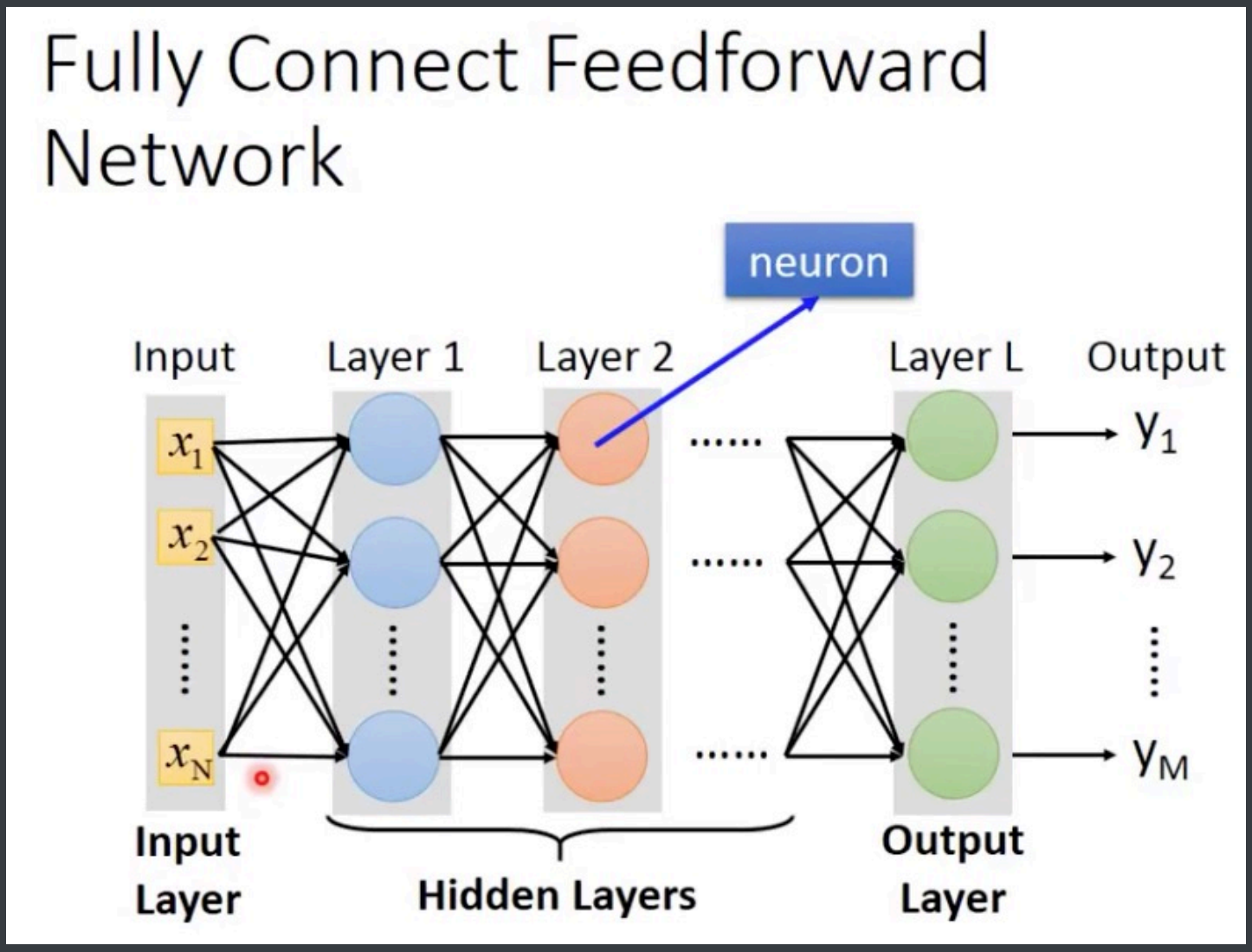
Continuing ...



Input is a vector (1, -1), output is a vector (0.62, 0.83)

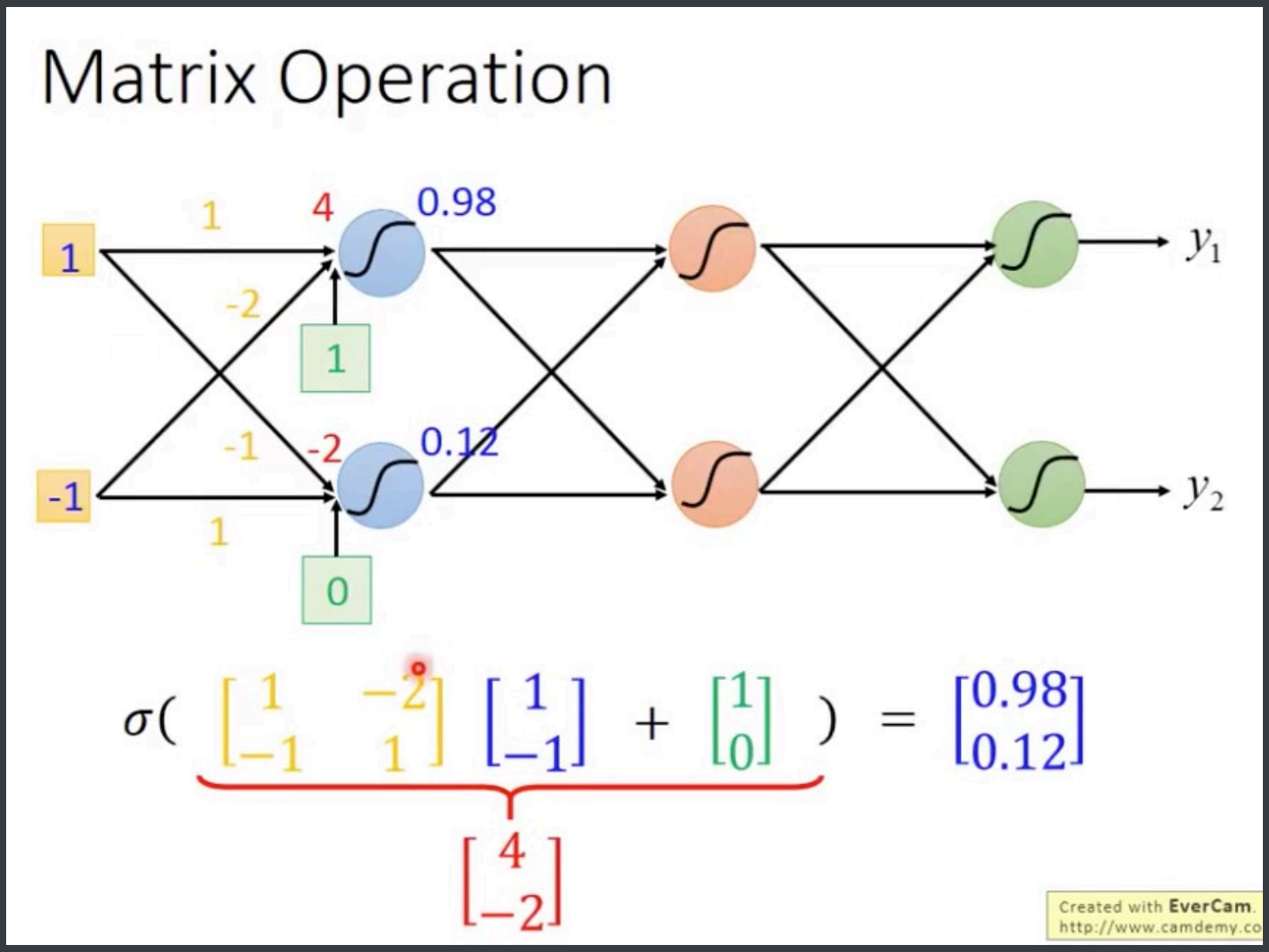
Given network structure, define a function set

Types of connections



Deep = many hidden layers

Matrix Operation



Feedforward Network Calculation Process:

(input multiply by matrix (weight) + bias) >> into sigmoid (any activitation function) >>> output