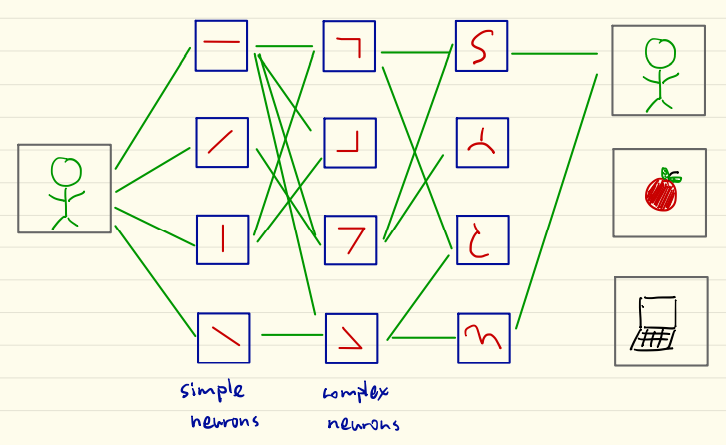
Part IV Deep Learning

˙Deep learning is a branch of ML and is based on neural networks (NN), which tries to mimic how human neurons work.

˙Deep learning has revolutionized computer vision and natural language processing and has been a critical factor behind many recent breakthroughs in AI & ML.

˙A brief history of deep learning:

-1959 Wiesel & Hubel: How visual information is proceeded in mammalian cerebral cortex.



˙1979 Fukushima proposed neocognitron based on Wiesel and Hubel’s discovery about primary visual cortex hierarchy. Capable of identifying handwritten characters.

˙1998 Lecun and Bengio proposed LeNet-5 which practically implements backpropagation.

Data Set 來源：

˙MNIST: 60000 training images + 10000 test images of human handwriting digits (adversarial, 因為是有挑過特別難辨識的)

˙Image Net: Tens of millions of labelled photographs

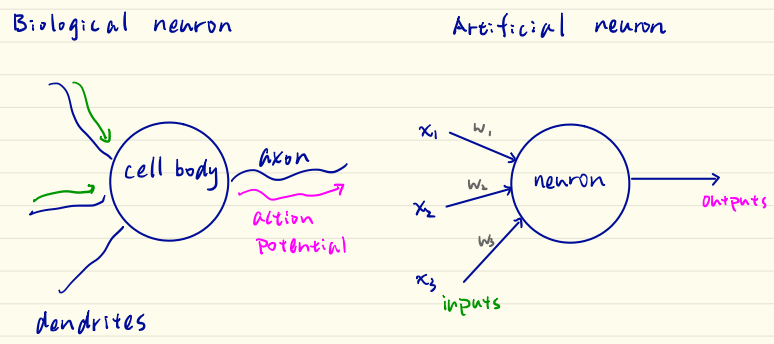
˙ILSVRC: Open challenge on a subset of Image Net

˙CIFAR

一開始的Deep Learning是Deep Neural Networks

Artificial Neuron

Perceptron (Rosenblatt 1958) 泛指halfspace



軸突

樹突

生物上的神經元：樹突接收刺激，如果訊號夠強(超過threshold)，則由軸突往下傳到下一層細胞。

Inputs: x1, x2, …, xn. Let x = [x1, x2, …, xn]T

Weights: w1, w2, …, wn. Let w = [w1, w2, …, wn]T

Threshold: -b where b is called bias

Perceptron algorithm:

=>

Ex: NCTUer detector. Threshold -b = 4

x1: knowledgeable x2: Handsome/Beautiful x3:19~25 yr old

w1 = 3 w2 = 2 w3 = 6

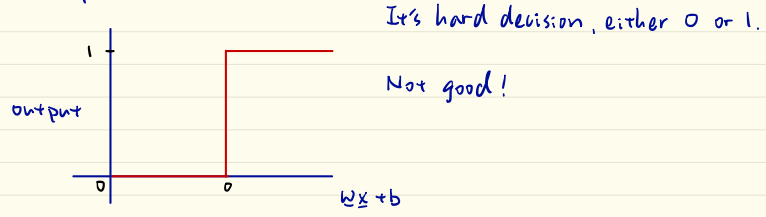
Person A: knowledgeable + H/B + age 30+ =>

Person B: NOT knowledgeable + H/B + age 40+ =>

=> output 1

=> output 0

Perceptron transition



(因為它只是中間產物，不想在這裡就做決定(hard decision)，想要保有soft information)

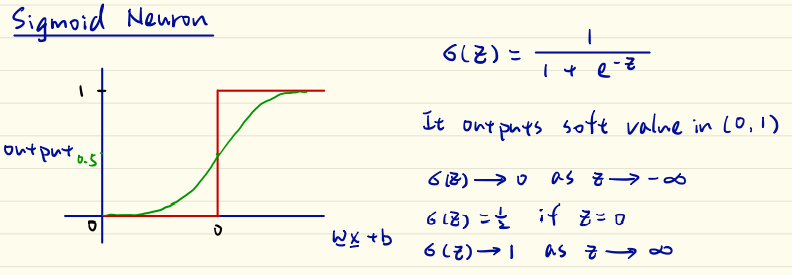
˙Do not distinguish very large and moderately large values.

˙Sharp transition around z. Discontinuity prevents usage of SGD.

: activation function

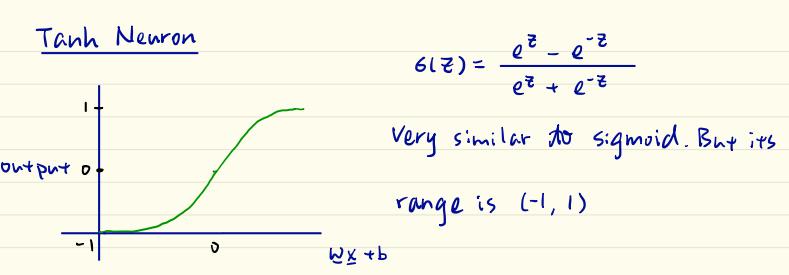
Perceptron is the first activation function, there are more.

Sigmoid Neuron



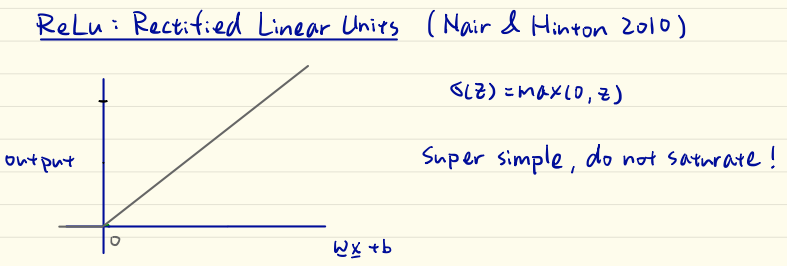
(通常最後一層用Sigmoid)

Tanh Neuron



Issue in Sigmoid & Tanh: They saturate! (在中間段可以很好的分別，但在兩端區別就很小(saturation))(這就是為甚麼作業2要做normalization，因為有可能全落在很大或很小的區段，造成無法分辨)(至於要用怎麼選擇normalization的方法，又是另一個問題)

ReLu: Rectified Linear Units (Nair & Hinton 2010) 目前最建議選的actuation func.



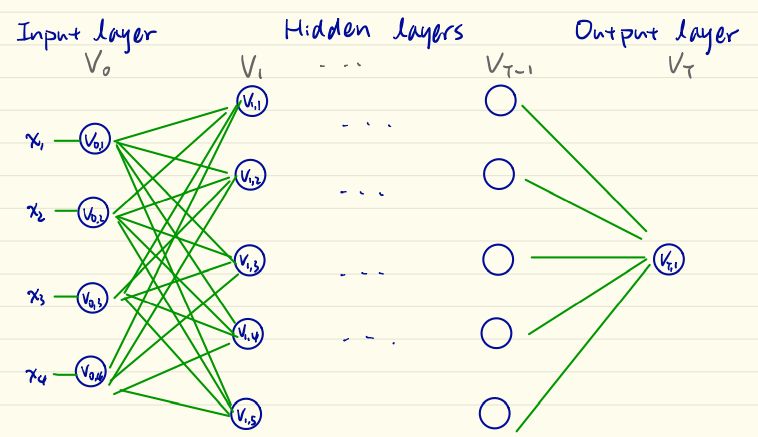
˙It’s piecewise linear: if and if

˙It was adopted in 2012 AlexNet.

˙Nowadays, it is the most widely used neuron for hidden layers.

˙How to choose neurons properly? It’s an art!

Artificial Neural Networks (ANN)



(使用者不用知道，黑盒子)

VT: Output layer

T: Depth of the network (不含最後一個 - output layer)

Deep learning (DNN) if (-> -> 至少2個hidden layer才算)

What to choose?

How many layers, neurons, width, input形式?

Train What? Train weight.

˙最後一個layer必為dense layer(每一個neuron都有連到output的edge，否則表示output根本沒參考到該neuron -> 該neuron沒有存在的意義)

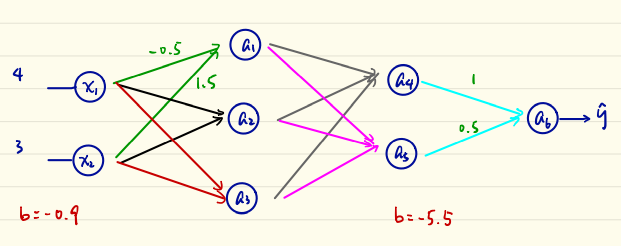
˙output layer也可以有很多neuron(用一樣的訊息可能判斷多個結果)

Associate with each edge is a weight

Given (x1, x2, x3, x4), there is a function g(x1, …, x4) that outputs the true label. ANN is a function approximator of g(.).

Ex: NCTUer detector

x1: knowledgeable x2: Handsome/Beautiful



Two dense hidden layers with ReLu neurons

Suppose we have done calculation and obtained

The output layer is Sigmoid:

This indicated that the neural net estimates that there is an 11.92% chance that this person is a NCTUer.

This procedure is called forward propagation.