Deep Learning Summary

Definition: ***Deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input.***

1. Key algorithm: Stochastic gradient descent (SGD) to find the best weights x.

   Weights $x = (A_1, b_1, \dots, A_L, b_L)$, where A_k, b_k are the weight and bias of every parameters in k layer in the neural network.

   SGD finds weights x that generalize-weights that will succeed on unseen vectors v from a similar population by minimizing loss function. A favorite loss function is ***Cross-entropy loss***.

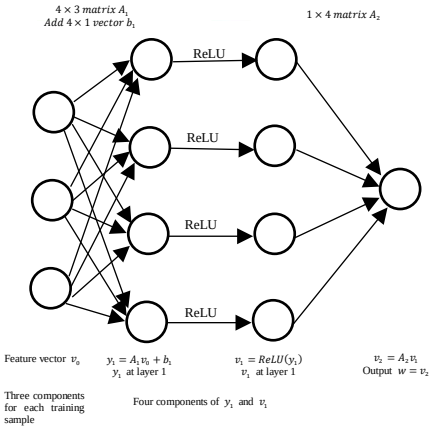
2. Key Nonlinearity: Ramp function $ReLU(y) = max(y,\ 0)$

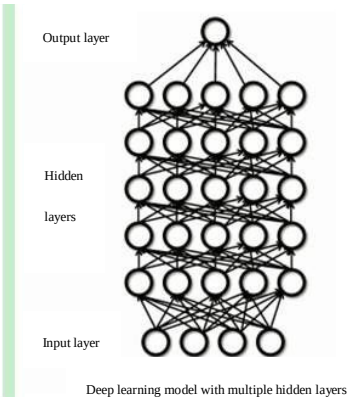
   Identify function. Control the activation of the next neuron.

3. Key operation: Composition $F = F_3(F_2(F_1(x, v)))$

   F_k is a piecewise linear function of a vector v_{k-1} of length N_{k-1}, and output a vector v_k of length N_k ready for input to F_{k+1}.

$$v_k = F_k(v_{k-1}) = ReLU(A_k v_{k-1} + b_k)$$





Deep learning model with multiple hidden layers

4. Key rule: Chain rule and Backpropagation

   Backpropagation is a method to compute derivatives quickly, using the chain rule. The goal is to find the right weights of which loss function depending on.

Application: Convolutional Neural Nets (CNN)

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of a series of convolutional layers that convolve with a multiplication or other dot product. The activation function is commonly a ReLU layer, and is subsequently followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution. The final convolution, in turn, often involves backpropagation in order to more accurately weight the end product.