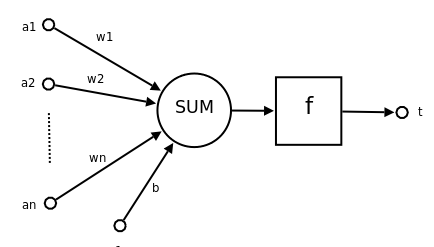
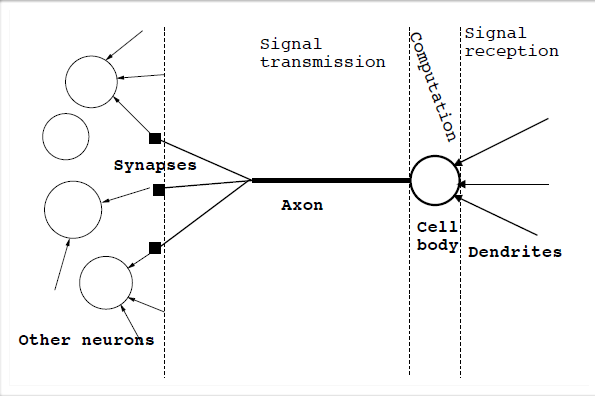
Neural Network簡介

Neural Network(NN)或(Artificial Neural Network)ANN，是一種模仿[人類神經網絡](https://zh.wikipedia.org/wiki/%E7%94%9F%E7%89%A9%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C)的結構的模型。使用人工神經元聯結進行計算，神經網絡是由大量的[神經](https://zh.wikipedia.org/wiki/%E7%A5%9E%E7%B6%93%E5%85%83)單元和之間相互聯接構成。每個神經元可以接受輸入值，並用一個輸出函數（activation function），來決定輸出，而任兩個單元間的連接代表信號的加權值（weight）。整個架構的輸出值依網絡的連接方式，隨著網絡中的參數weight、bias和activation function的不同而有著不同的學習能力。

單一神經單元*(Wikipedia)*

單一神經元即具有邏輯分辨能力，以一個2維二元(0或1)輸入為例，可以藉由調整參數，來得到and,or等邏輯分辨，以or來說(0,0)不會激發，而這三種輸入(0,1),(1,0),(1,1)都可以激發神經元，and則是只有(1,1)會激發。

**Neural Network(NN)**的構想來自於生物神經，每一個神經細胞主體，都會接著樹突，用來接收輸入進來的神經訊號，以及軸突，輸出結果，給多個相鄰神經細胞的樹突，形成生物神經網絡，類神經單元即是模仿神經細胞，連接多個輸入神經元的訊號，不同神經元的訊號輸入都各自給它乘上weight和bias，表示著神經輸入訊號重要與否，接著用一個activation函數，來決定訊號是否輸出到下一層的神經單元，表示著總輸入值是否有超過一個門檻來使神經激發或者抑制

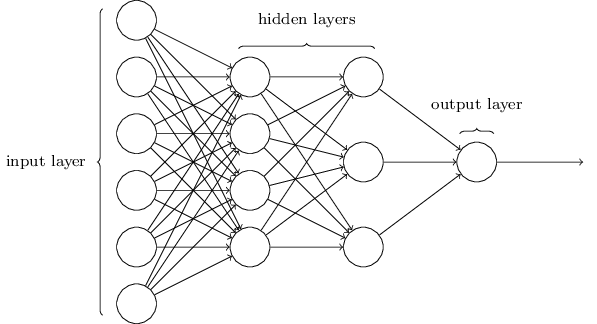
生物神經細胞

*(Hugo Larochelle Neural Networks Slides)*

Deep Neural Networks(DNN)

單一神經元的運算結果可以代表著神經元具備一些邏輯分辨的能力，若我們將多個神經原並排成一層(layer)，並且架構多層神經元，形成一個深度神經網絡，我們可以預期到這個深度神經網絡(Deep Neural Networks)，將具備處理更高維、更複雜任務的能力，例如：圖形辨識，語音辨識等等。

整個類神經架構，可以藉由監督或非監督訓練每層神經單元到下層單元之間的weight和bias，藉由調整參數，得到較符合測試情境的架構，測試時，在給定一串輸入訊號時，用經過加權後的分數來決定下一層每個神經單元是激發(activation)還是抑制(inhibition)，以次由下而上，最後決定輸出。



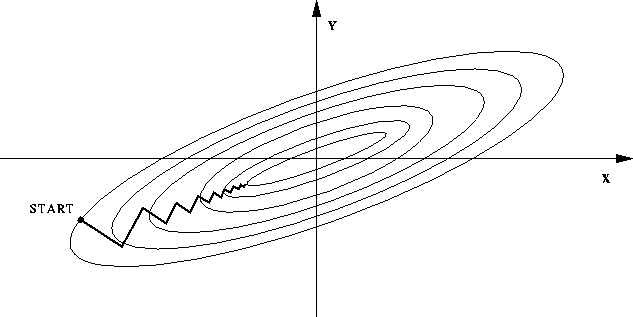
*(http://neuralnetworksanddeeplearning.com/chap1.html)*

如何訓練DNN中的參數：

”Deep Neural Networks for Acoustic Modeling in Speech Recognition ,Geoffrey Hinton,2012 “

Back Propogation:

一般而言，我們使用back propogation的方式，先隨機設定參數W和b，將訓練資料輸入，得到一個和期望差距很大的輸出，接著由後(輸出層)往前(輸入層)，調整各層的參數，使用的方法為stochastic gradient descent(SGD)，一開始，我們先定義Cost Function(C)，代表著期望輸出值和真實輸出值的分布差異，C越小代表差異越小，輸出越接近期望輸出值，接著想像各層之間的參數W或b，構成一個多維度的空間平面，空間中每個點代表一組W值，這組W對應到一組輸出，因此算出一個C，從一組隨機的W開始，每次訓練時都朝著C減少的方向(gradiant)移動調整W，直到收斂，此一方法即為stochastic gradient descent



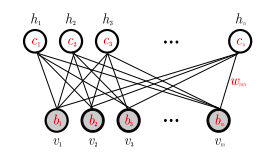
以二維W為例，從一點W開始每一步都朝著C減少最多的方向移動，直到local最小值收斂*(http://trond.hjorteland.com/thesis/img208.gif)*

然而使用SGD做back propogation的訓練，有一定的限制，例如說，起使位置一開始設得離目的地太遠，而每一步移動的距離越來越小，以至於要訓練非常多次才能達到，或是因為步距已經收斂而永遠無法到達最低點，因此我們需要一個好的起使位置來做back propogation，即pre-training

Restricted Boltzmann Machine(RBM)

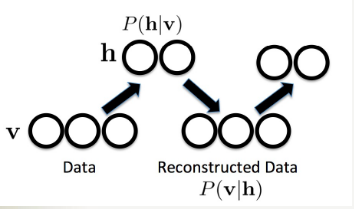
“An Introduction to Restricted Boltzmann Machines ,Asja Fischer and Christian Igel,2012”

“A Practical Guide to Training Restricted Boltzmann Machines ,Geoffrey Hinton,2010”

****

在訓練Neural Networks的參數時，RBM即是先將相鄰兩層視為一個undirected model，我們可以利用參數從可觀測層(visible layer)向上求得隱藏層(hidden layer)，也可以由隱藏層(hidden layer)向下求得可觀測層(visible layer)

接著，定義一參數E(Energy)代表著visible(v)和hidden(h)層之間所有參數狀態，我們想要使得此一模型在某一參數狀態下，得到v機率最高，Hinton等人發展出此一演算法，使用Gibbs Sampling，反覆的利用參數求v和h，近一步調整參數W，或者，以Contrast Divergence(CD-1)方式，僅做一次取樣(reconstructed data)，求得W變化量，以同樣方法訓練完v和h間的所有參數，在此方法中，訓練好的參數，即可作為backpropogation一個好的起始點



*(R. Salakhutdinov.*[*Deep Learning Tutorial*](h7p://www.utstat.toronto.edu/~rsalakhu/isbi.html)*)*

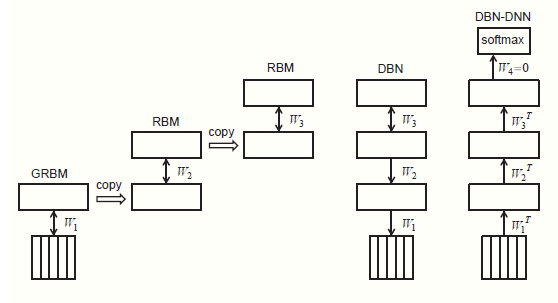
Deep Belief Networks and Autoencoders

“Deep Neural Networks for Acoustic Modeling in Speech Recognition ,Geoffrey Hinton,2012 “

“Acoustic Models Using Deep Belief Networks, [A Mohamed](https://scholar.google.com.tw/citations?user=tJ_PrzgAAAAJ&hl=zh-TW&oi=sra), 2012”

“Reducing the Dimensionality of Data with Neural Networks,Science,G Hinton,2006”

RBM的架構訓練可觀測層v和隱藏層h之間的參數，若把多層的RBM架構疊在一起，將上一個的RBM的隱藏層輸出，作為下一個RBM的可觀測層輸入，形成一個Deep Belief Networks，可以利用RBM由下而上訓練參數，同時，也可以由上而下，從輸出層求得原始輸入層



(Deep Neural Networks for Acoustic Modeling in Speech Recognition ,Geoffrey Hinton,2012)

使用DBN架構，我們可以將高維度的資料作為底層的輸入，並且逐一降低各層神經元數目，使用RBM向上訓練，可以得到一個較低維度的資料，而這個低維度資料也可以利用DBN向下求得原始資料，亦即此系統將高維資料encode成低維度、高階的抽象概念，也可以將此低維度資料decode成原始高維資料，此概念就是”Autoencoder”，而在此高階空間中，我們可以更進一步對原始資料做歸類classification，例如辨識手寫數字資料，如下圖：



圖為將數字圖像作為autoencoder的輸入，使用RBM訓練參數，輸出結果比較