

程式人《十分鐘系列》



用十分鐘理解 神經網路發展史

陳鍾誠

2016 年 11 月 23 日

我大學念交大資訊科學系的時候

- 聽到《類神經網路》這個詞

就覺得超酷的！

於是跑去修這門課

- 結果第一堂課，老師就播放一個日本人將《狗鼻子的神經細胞》鋪在電路上，用《神經網路》訓練後進行《味道辨識》的例子

然後老師要我們去看一篇論文

- 並實作出一個手寫數字辨識系統。

我花了很多時間寫

- 最後終於根據那篇論文做出一個
手寫數字辨識系統
- 效果還算不錯！

這是我第一次

- 和神經網路交手的經驗！

問題是、程式寫完之後

- 我對這些東西也就忘光了！

只記得

- 裡面的反傳遞演算法，有用到微積分
- 而且還是偏微分，還蠻困難的！

這是我第一次

- 發現寫程式竟然要用到這麼難的數學！

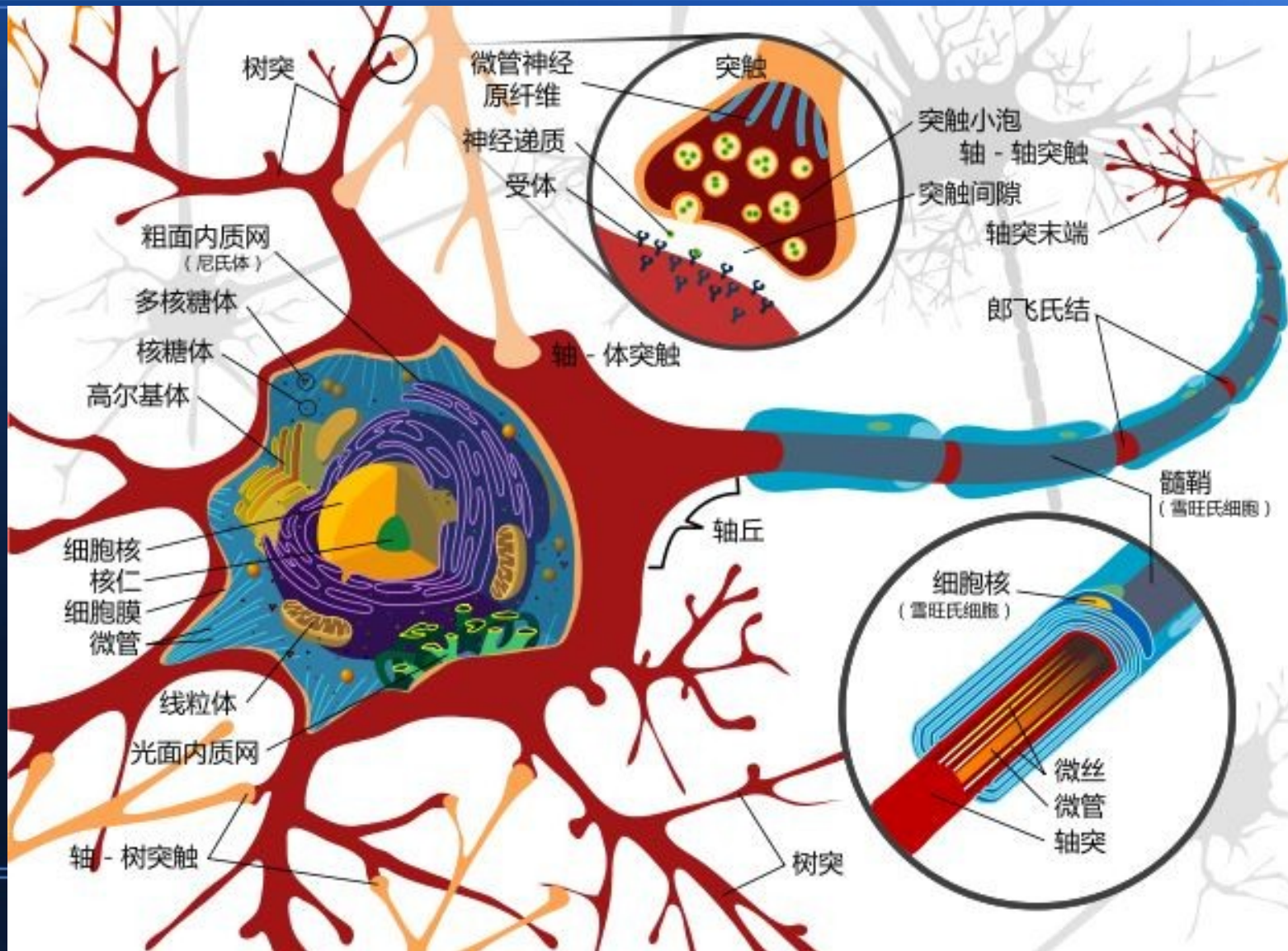
否則我還真的不知道

- 到底學微積分是要幹嘛用的！

現在、就讓我們來看看

- 甚麼是神經網路？
 - 微積分如何用在神經網路上
- 等問題吧！

生物學上的神經網路 長得像這個樣子！



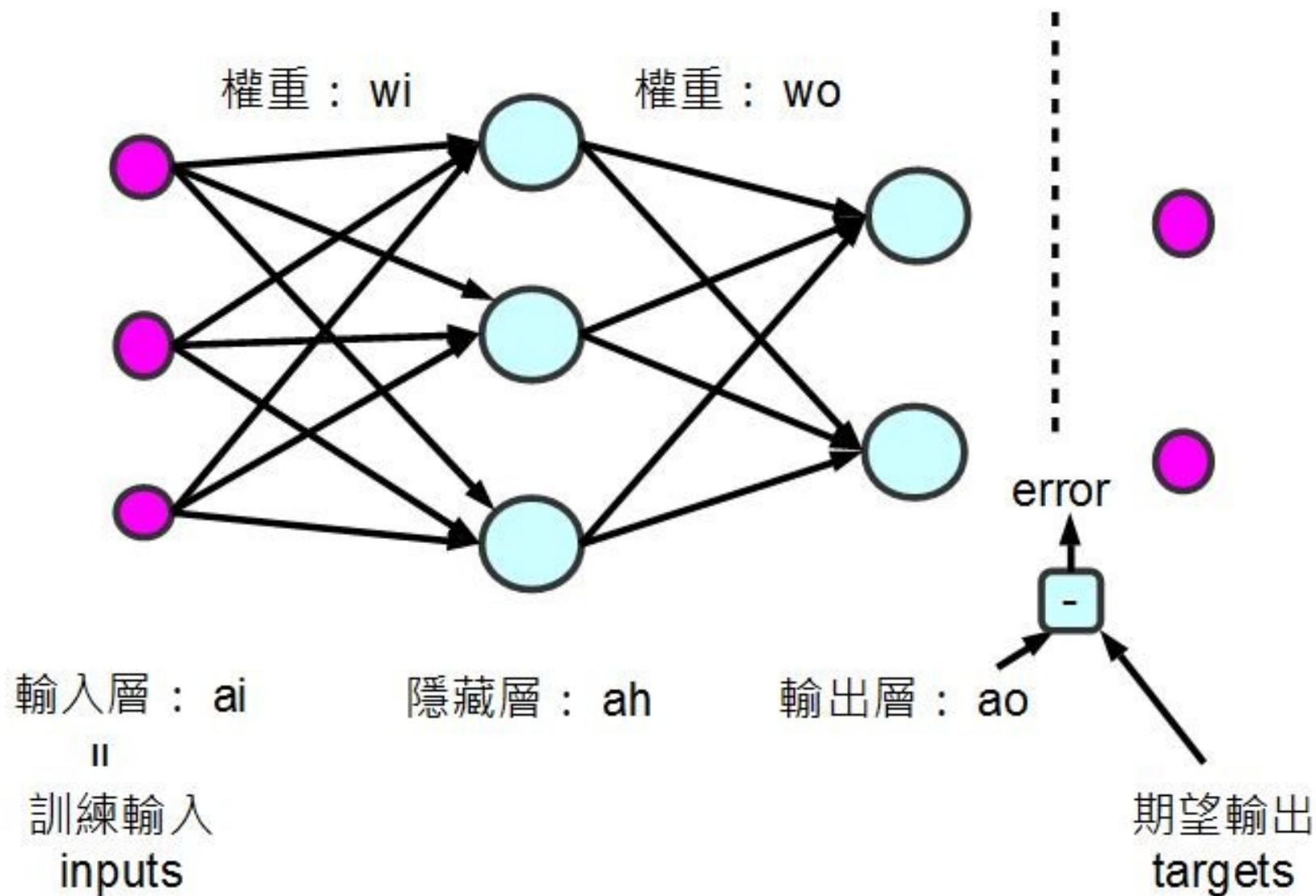
裡面有很多東西、包含

- 神經細胞、樹突、軸突
- 甚至還有《細胞核、細胞質、細胞膜、高基氏體、 …》

但是程式中的神經網路

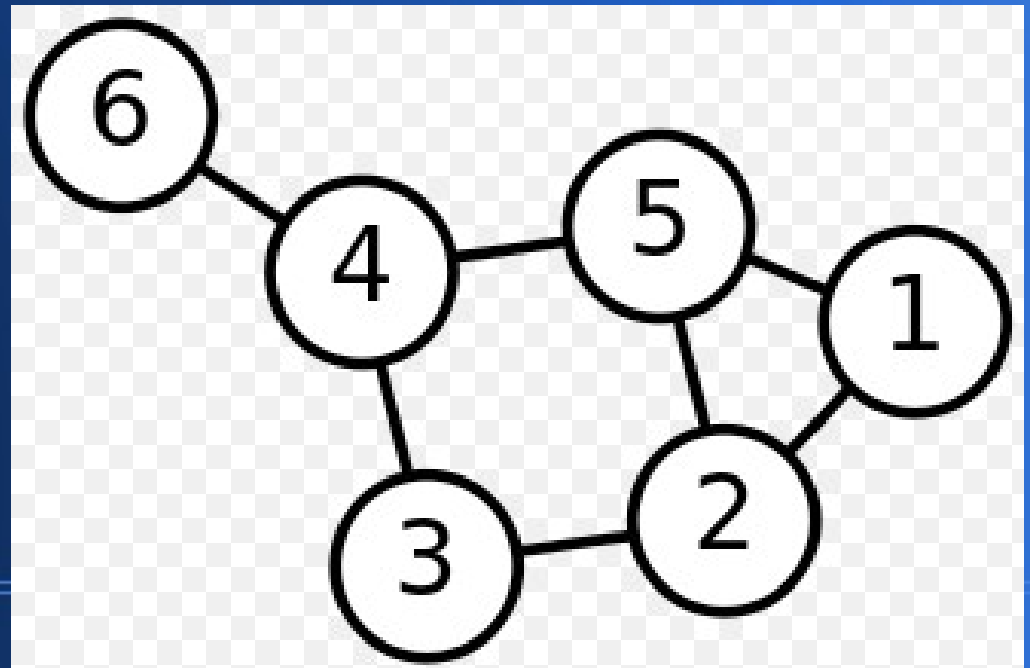
- 比較簡單一些！

通常長得像這個樣子



如果您念過圖形理論

- 會發現神經網路就是一個《圖形》
- 包含《節點》 (node) 和《連線》 (edge)
- 只不過連線上有些權重而已！



問題是

- 這樣的圖形，怎麼用來模仿大腦的神經系統，讓電腦能夠擁有智慧呢？

想要理解這件事

- 我們必須先回顧一下

《神經網路》的發展歷史！

西元 1872 年

- 義大利的醫學院畢業生高基 (Camillo Golgi)，在一次意外中，將腦塊掉落在硝酸銀溶液中。數週後，他以顯微鏡觀察此腦塊，成就了神經科學史上重大里程碑——「首次以肉眼看到神經細胞」。
- 於是 Golgi 發明了鉻酸銀染色法，讓生物學家可以清楚觀察到神經纖維走向。

Camillo Golgi
卡米洛·高基



出生	1843年7月7日 義大利布雷西亞
逝世	1926年1月21日 義大利帕維亞
研究領域	解剖學家、病理學家
任職於	帕維亞大學 博洛尼亞大學 錫耶納大學
母校	帕維亞大學
獲獎	 諾貝爾生理學或醫學獎 (1906年)

後來在 1898 年

- 高基又發現了

《高基氏體》

Golgi apparatus

卡米洛·高基（義大利語：**Camillo Golgi**，1843年7月7日－1926年1月21日），是一位義大利醫師與科學家，高基氏體的發現者，出生於布雷西亞。1906年因為神經系統的研究，而與西班牙的桑地牙哥·拉蒙卡哈共同獲得諾貝爾生理學或醫學獎。

於1898年發現高基氏體，主要功能為處理細胞膜、溶體或內體與細胞生產的蛋白質，將它們分到不同的小泡，是細胞的中心傳送系統。

可惜的是

- Golgi 雖然發明了鉻酸銀染色法，讓生物學家可以清楚觀察到神經纖維走向。
- 但是他本人卻因為支持 Joseph von Gerlach 提出的《網狀理論》(Reticular theory)，因而錯失了進一步開創《神經元理論》的契機！

還好

- Ramón Cajal 透過這個技術觀察神經細胞，並採用《神經元學說》neuron doctrine 解釋神經的作用，因而開創了《神經科學》。
- 後來他被認為是神經科學之父。

Santiago Ramón y Cajal

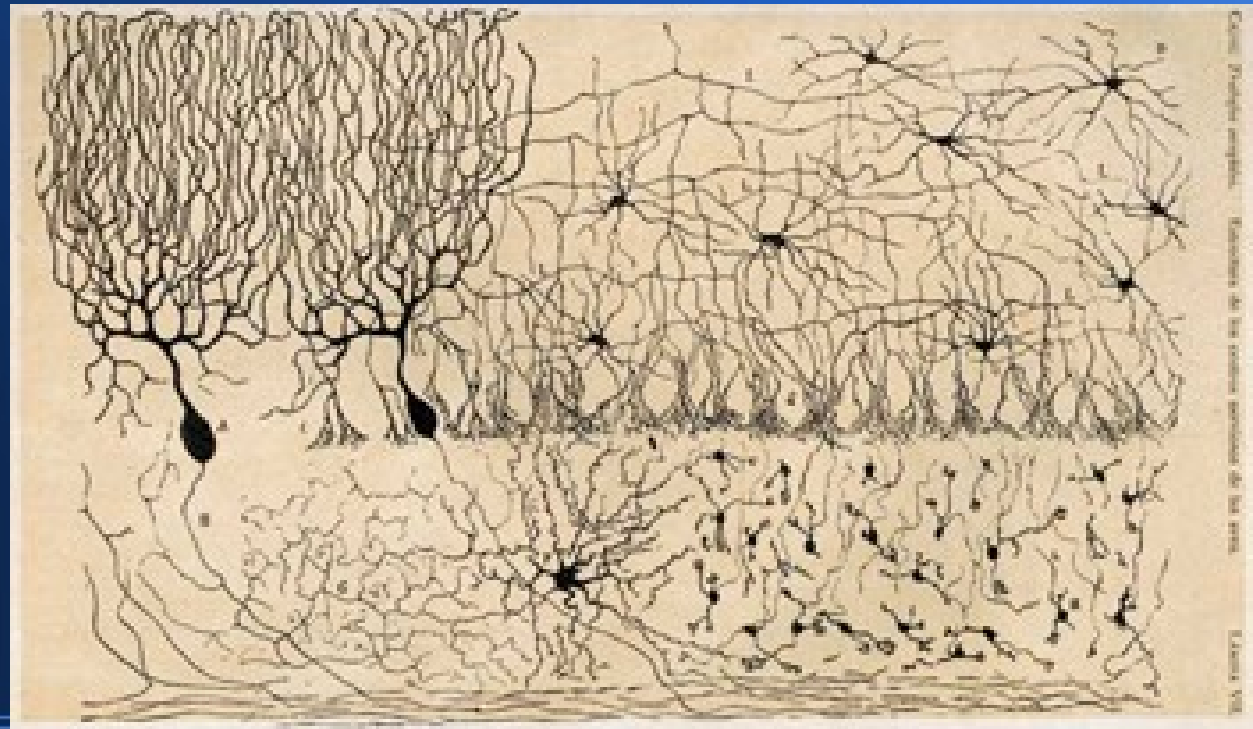
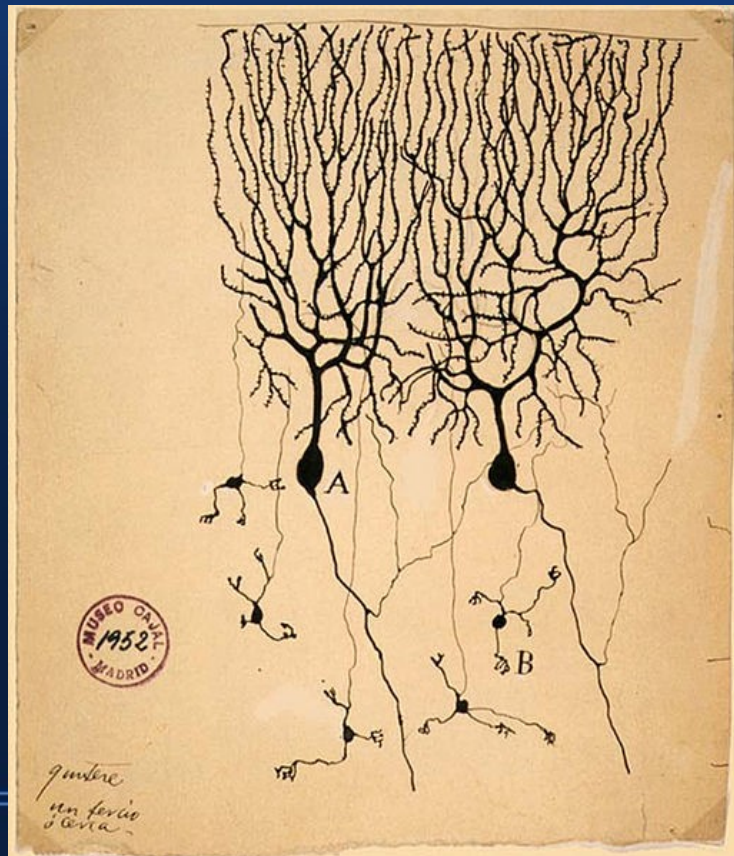


S. Ramón Cajal

Born	1 May 1852 Petilla de Aragón, Navarre, Spain
Died	18 October 1934 (aged 82) Madrid, Community of Madrid, Spain
Nationality	Spain
Fields	Neuroscience
Known for	the father of modern neuroscience

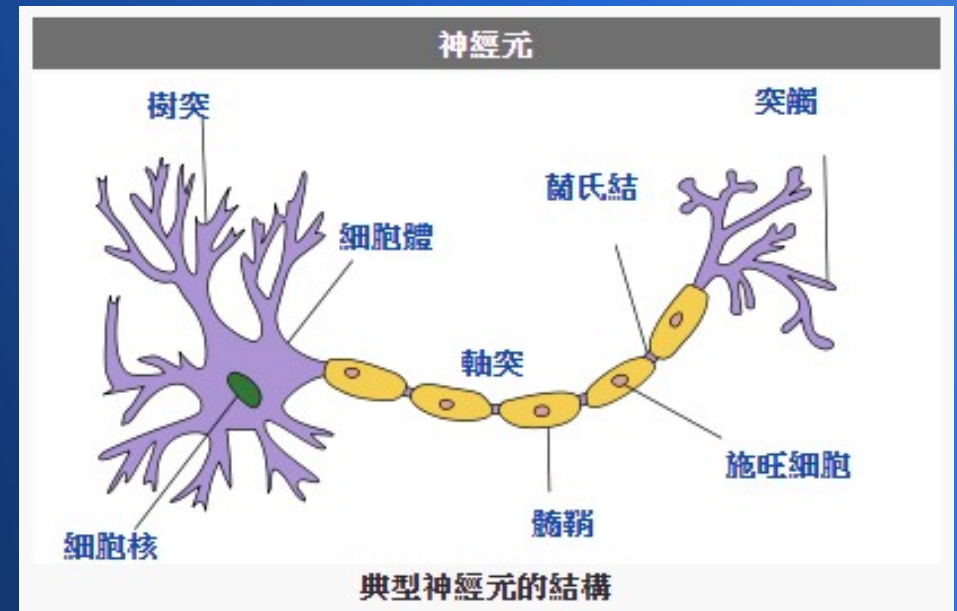
Ramón Cajal 很會畫畫

- 他把看到的神經元結構都畫下來



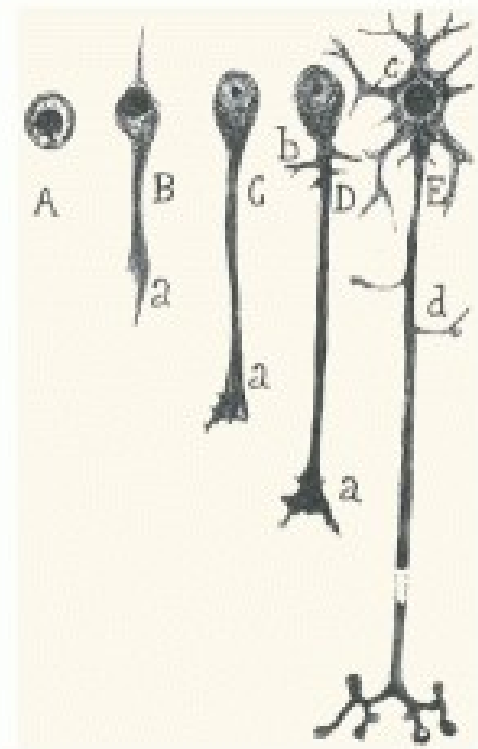
Ramón Cajal 發現

- 神經元細胞通常具有方向性，而且軸突 (axon) 通常結束於灰質 (gray matter)
- 因此認為神經傳導的方向是《神經元 => 軸突 => 樹突》



另外 Ramón Cajal 又發現

- 細胞的軸突和樹突是慢慢從細胞體生長出來的，並且發現了引導軸突生長方向的生長錐 (growth cone) 之存在



Cajal's painting – the development of a mature neuron

後來在 1906 年

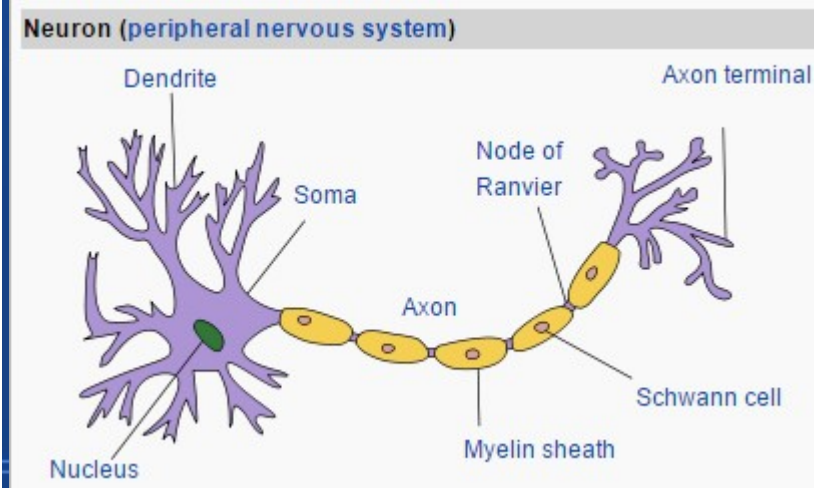
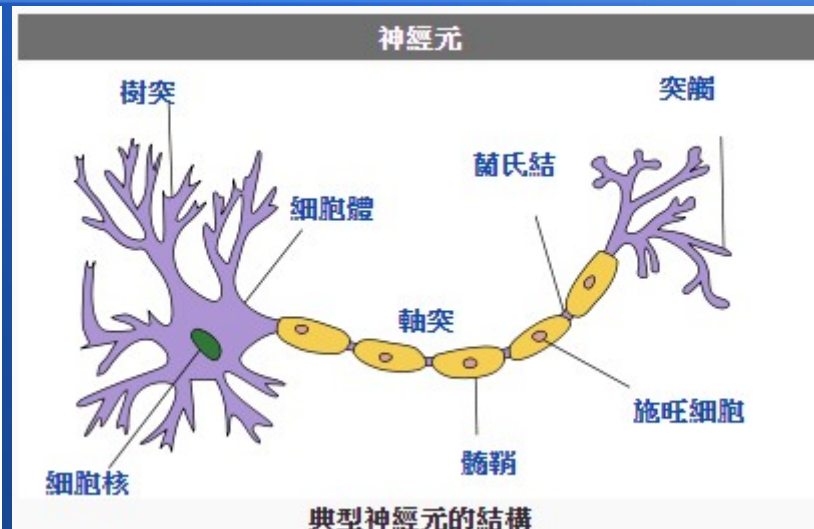
- 這兩人因此得了諾貝爾生理醫學獎！

於是神經的結構與功能越來越清楚

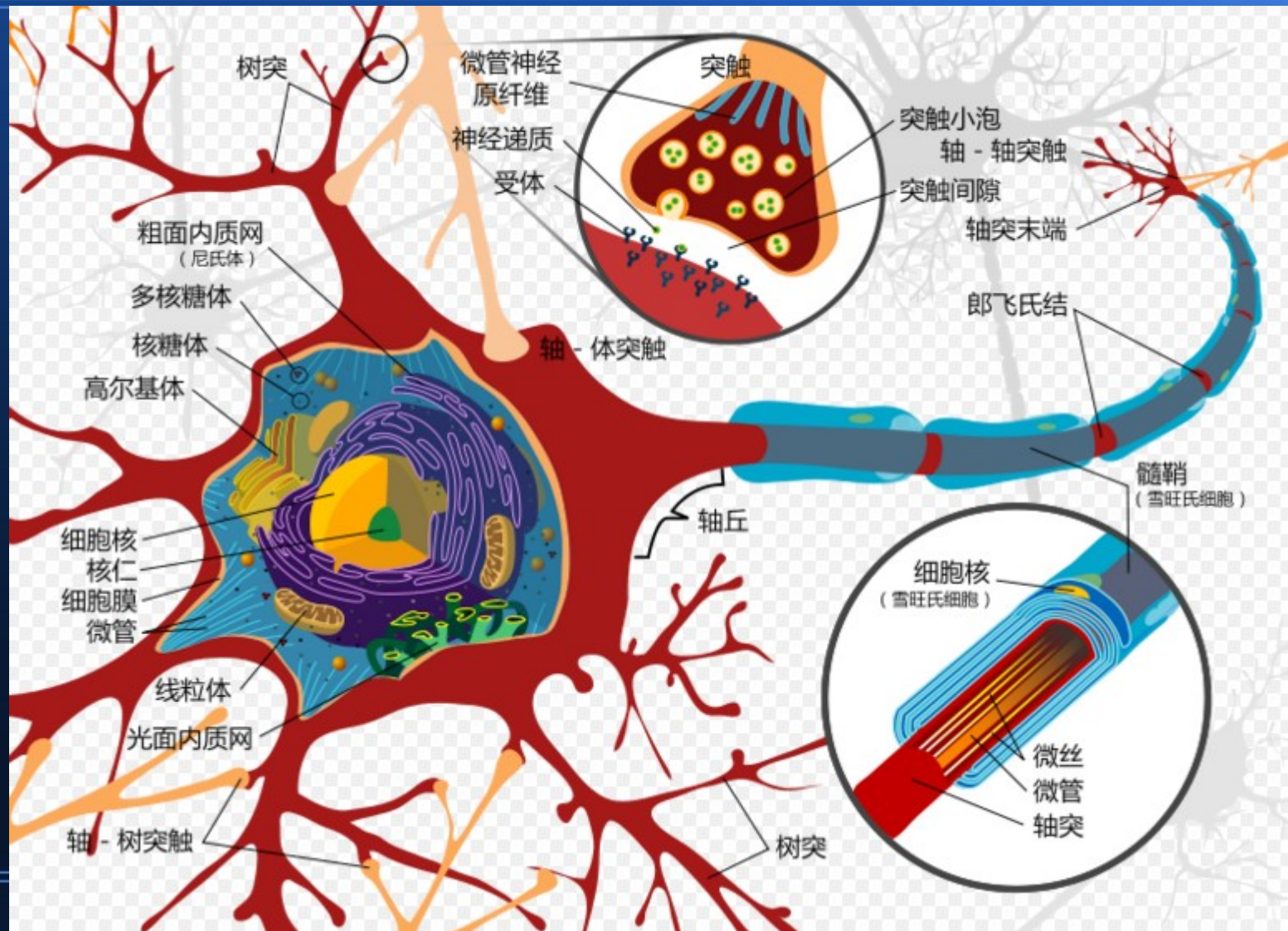
雖然神經元形態與功能多種多樣，但結構上大致都可分成細胞體（cell body, or soma）和神經突兩部分。神經突又分樹突（Dendrites）和軸突（Axon）兩種。軸突往往很長，由細胞的軸丘（axon hillock）分出，其直徑均勻，開始一段稱為始段，離開細胞體若干距離後始獲得髓鞘，成為神經纖維。習慣上把神經纖維分為有髓纖維與無髓纖維兩種，實際上所謂無髓纖維也有一薄層髓鞘，並非完全無髓鞘。

胞體的大小差異很大，小的直徑僅5~6 μm ，大的可達100 μm 以上。突起的形態、數量和長短也很不相同。樹突多呈樹狀分支，它可接受刺激並將衝動傳向胞體；軸突呈細索狀，末端常有分支，稱軸突終末（axon terminal），軸突將衝動從胞體傳向終末。通常一個神經元有一個至多個樹突，但軸突只有一條。神經元的胞體越大，其軸突越長。

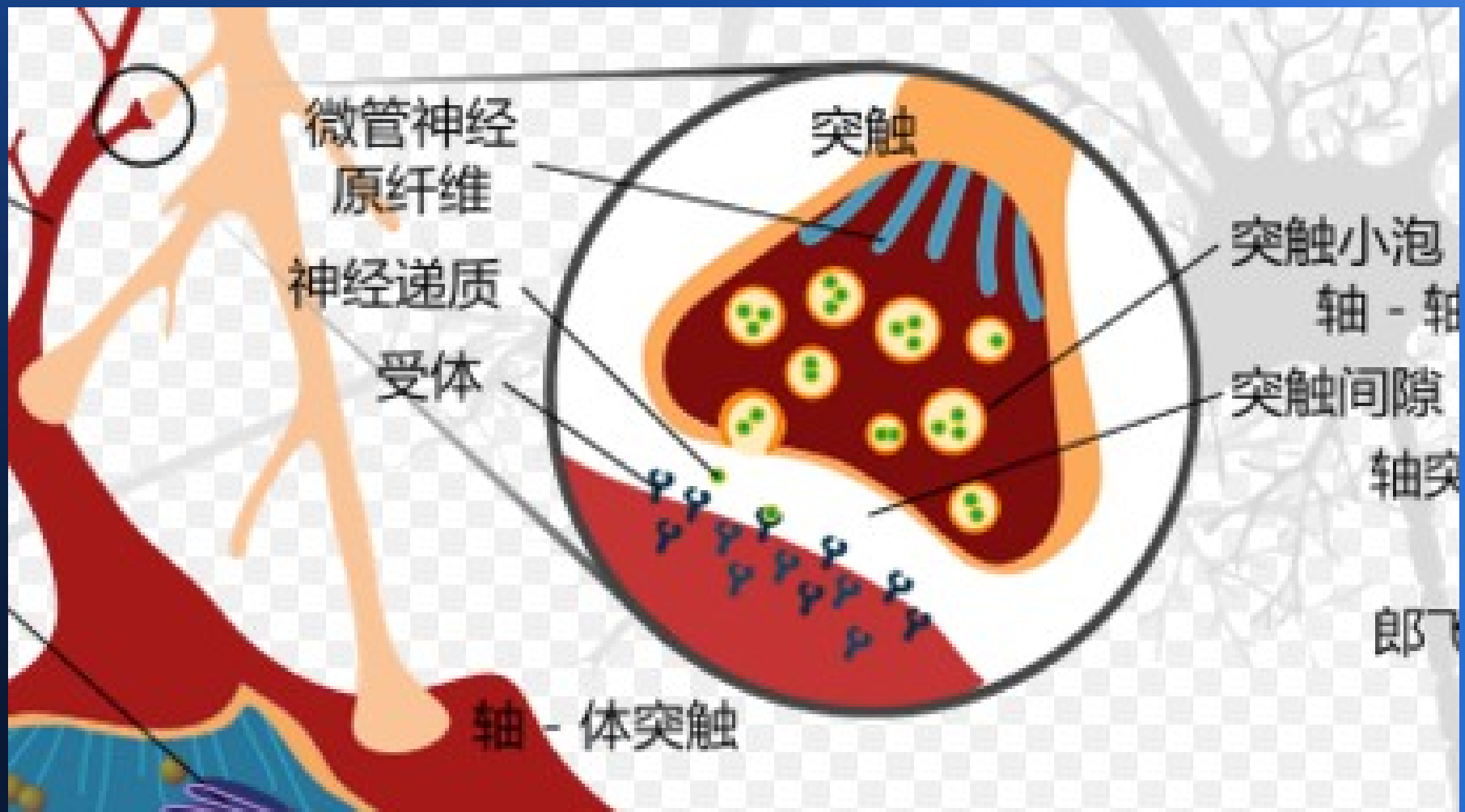
不論是何種神經元，皆可分成：接收區（receptive zone）、觸發區（trigger zone）、傳導區（conducting zone），和輸出區（output zone）。



後來的研究者更詳細的 揭露了神經元的結構

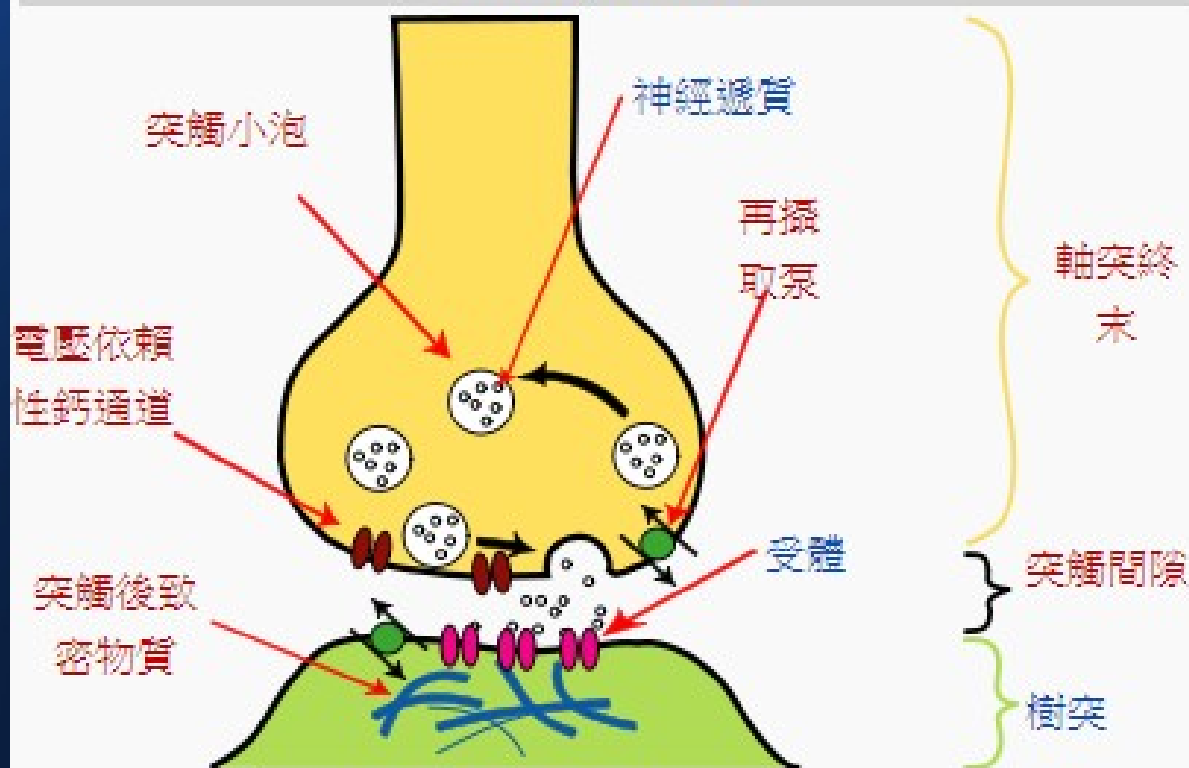


其中兩個神經元間接觸的點
稱為突觸 (synapse)

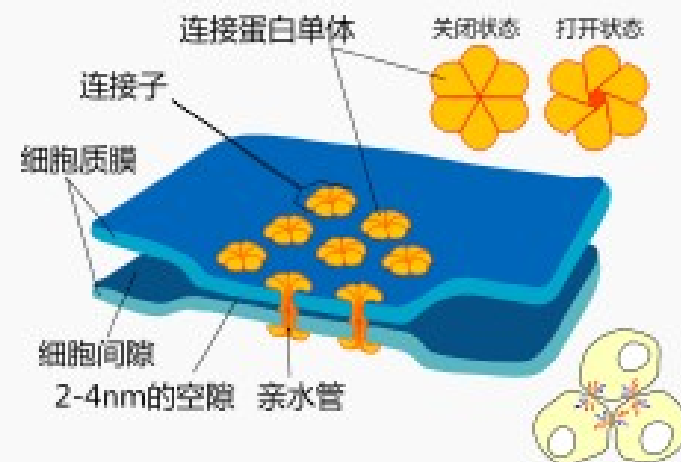


這些突觸上會有化學反應與電反應

典型突觸的結構



電突觸



電突觸的間隙連接

識別標示

TH H1.00.01.1.02024 [🔗](#)
FMA FMA:67130 [🔗](#)

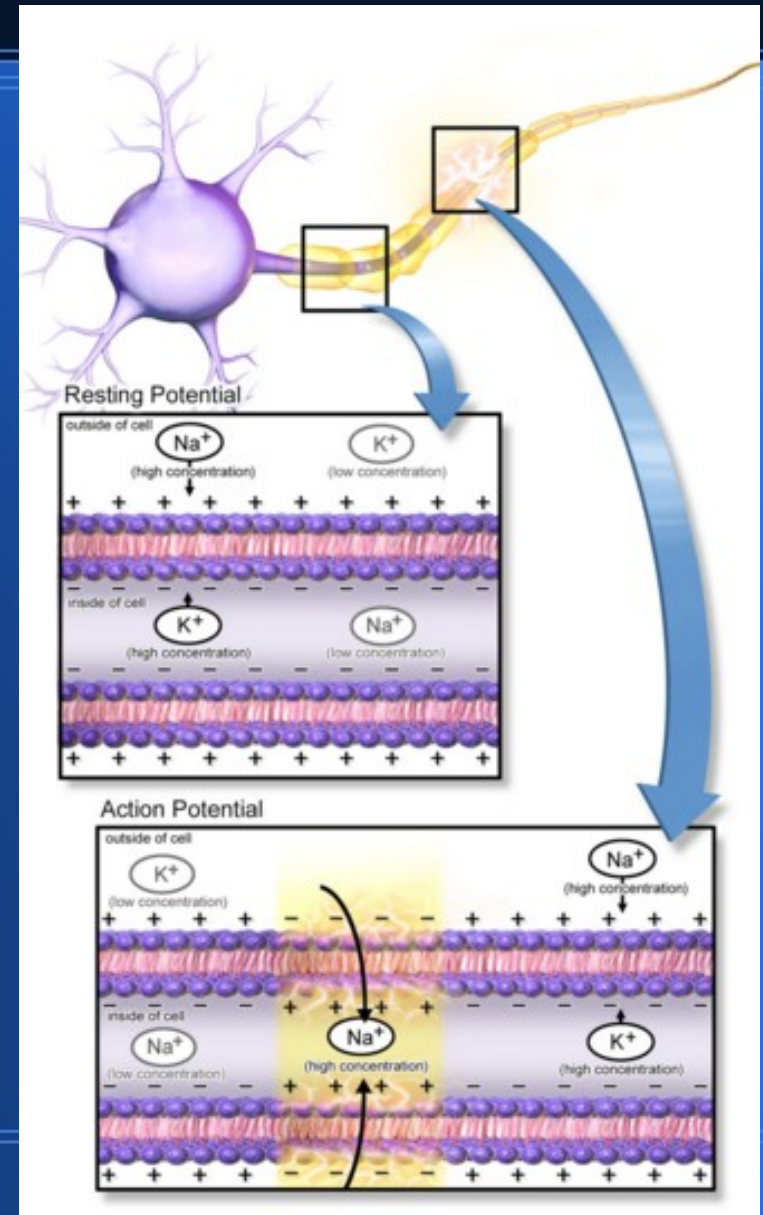
電突觸的反應比化學突觸快

電突觸是神經元之間**突觸**的一種，是以直接電氣方式耦合。電突觸是以兩個神經元之間相距僅**2至4奈米**的**縫隙連接**作為傳遞信號的地方^[1]，相對於**20-40奈米**的**化學突觸**來說小得多，帶電的**離子**可以在此交換，並引起神經衝動。對於到多數動物，同時存在電突觸和化學突觸這兩種神經突觸。

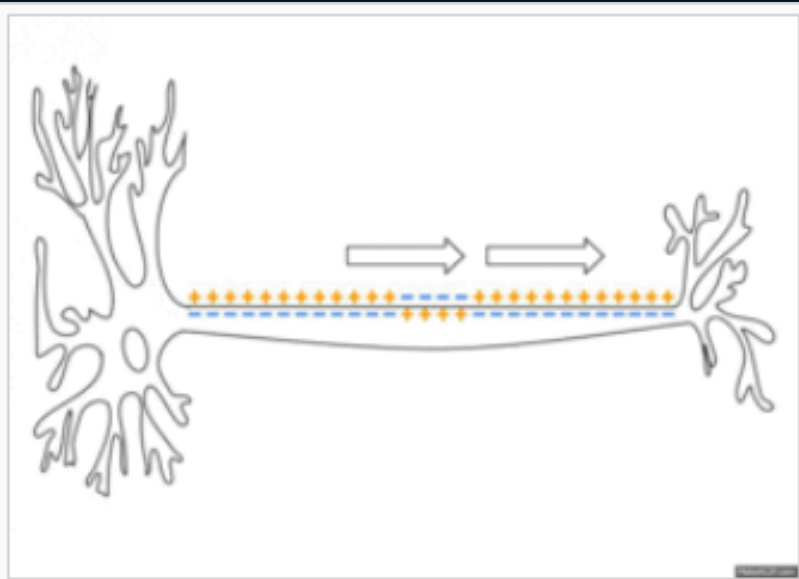
電突觸的優點是傳播速度較快，可以在更短的時間內傳遞信息。但是，由於缺乏**增益**，這種突觸的突觸後神經元受到的刺激的程度總是和突觸前神經元的相同或更小。故電突觸常常存在於神經系統中需要快速回應的地方，比如防衛反射。電突觸的另一種特點是，一般情況下它們是雙向的（即允許興奮從突觸兩側的任意細胞發出），但有些電突觸仍是單向的。

這些電訊號經由軸突和樹突傳導

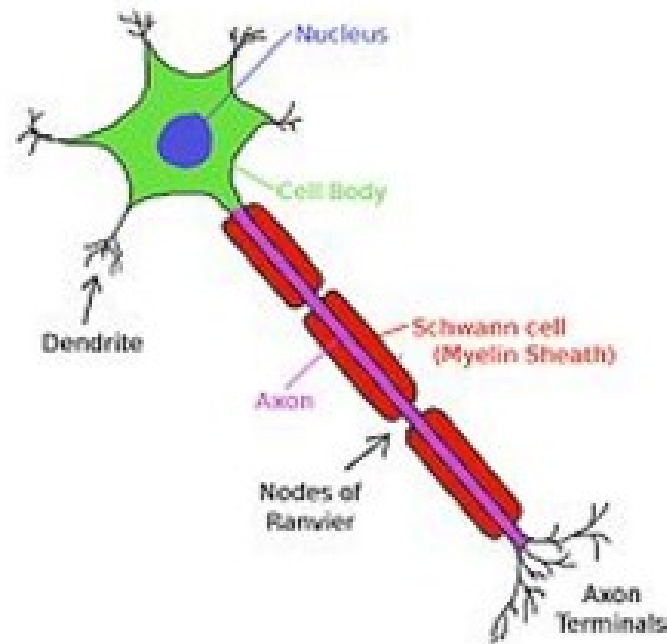
- 其中化學傳導
是由鈉鉀等離子
所導引的！



這些電衝動是透過神經軸傳導的



As an action potential travels down the **axon**, there is a change in polarity across the **membrane**. The Na^+ and K^+ gated **ion channels** open and close as the membrane reaches the **threshold potential**, in response to a signal from another **neuron**. At the beginning of the action potential, the Na^+ channels open and Na^+ moves into the axon, causing **depolarization**. **Repolarization** occurs when the K^+ channels open and K^+ moves out of the axon. This creates a change in polarity between the outside of the cell and the inside. The impulse travels down the axon in one direction only, to the **axon terminal** where it signals other neurons.



In **saltatory conduction**, an action potential at one **node of Ranvier** causes inwards currents that depolarize the membrane at the next node, provoking a new action potential there; the action potential appears to "hop" from node to node.



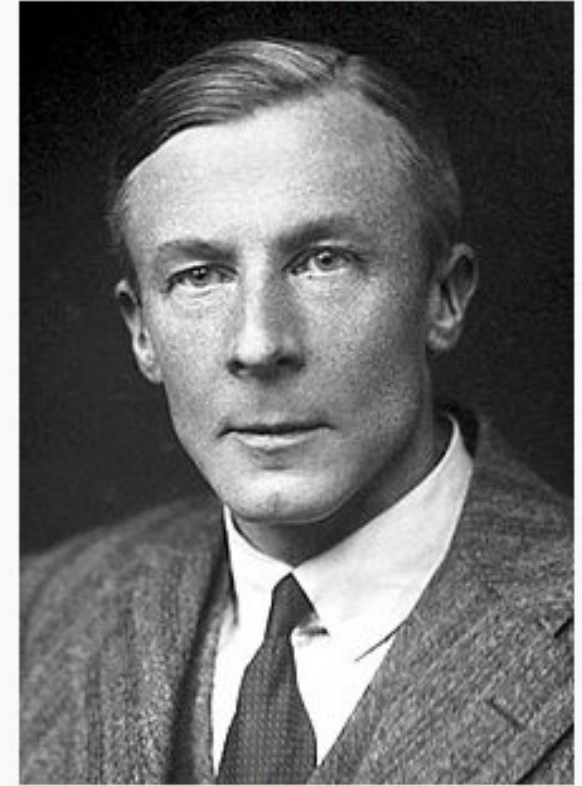
1928 年、Edgar Adrian

研究神經的電衝動

- 他用無線電技術中使用的放大器和毛细靜電計 (Capillary electrometer) 結合，記錄下非常微小的神經電位變化。
- 結果發現感受器刺激強度增加，會造成衝動頻率提高，參與衝動的神經纖維數量增加。

參考：神经脉冲——英国电生理学家阿德里安

阿德里安勋爵
Lord Adrian



出生	1889年11月30日 英国伦敦
逝世	1977年8月4日 英国剑桥郡剑桥
国籍	 英国
研究领域	电生理学
任职于	剑桥大学
母校	剑桥大学
获奖	 诺贝尔生理学或医学奖 (1932年) 科普利奖章 (1946年)

Edgar Adrian 是從《一隻蟾蜍》 身上發現這件事的

Continuing earlier studies of Keith Lucas, he used a capillary electrometer and cathode ray tube to amplify the signals produced by the nervous system and was able to record the electrical discharge of single nerve fibres under physical stimulus. An accidental discovery by Adrian in 1928 proved the presence of electricity within nerve cells. Adrian said,

"I had arranged electrodes on the optic nerve of a toad in connection with some experiments on the retina. The room was nearly dark and I was puzzled to hear repeated noises in the loudspeaker attached to the amplifier, noises indicating that a great deal of impulse activity was going on. It was not until I compared the noises with my own movements around the room that I realized I was in the field of vision of the toad's eye and that it was signaling what I was doing."

A key result, published in 1928, stated that the excitation of the skin under constant stimulus is initially strong but gradually decreases over time, whereas the sensory impulses passing along the nerves from the point of contact are constant in strength, yet are reduced in frequency over time, and the sensation in the brain diminishes as a result.



當 Edgar Adrian 把 蟾蜍視神經接到放大器後連接上喇叭

- 然後在很黑暗的房間裡，他意外地發現自己在蟾蜍面前走動竟然會造成喇叭出現噪音，後來他發現這些噪音是蟾蜍看到他的人影所造成的神經衝動訊號。

"I had arranged electrodes on the optic nerve of a toad in connection with some experiments on the retina. The room was nearly dark and I was puzzled to hear repeated noises in the loudspeaker attached to the amplifier, noises indicating that a great deal of impulse activity was going on. It was not until I compared the noises with my own movements around the room that I realized I was in the field of vision of the toad's eye and that it was signaling what I was doing."

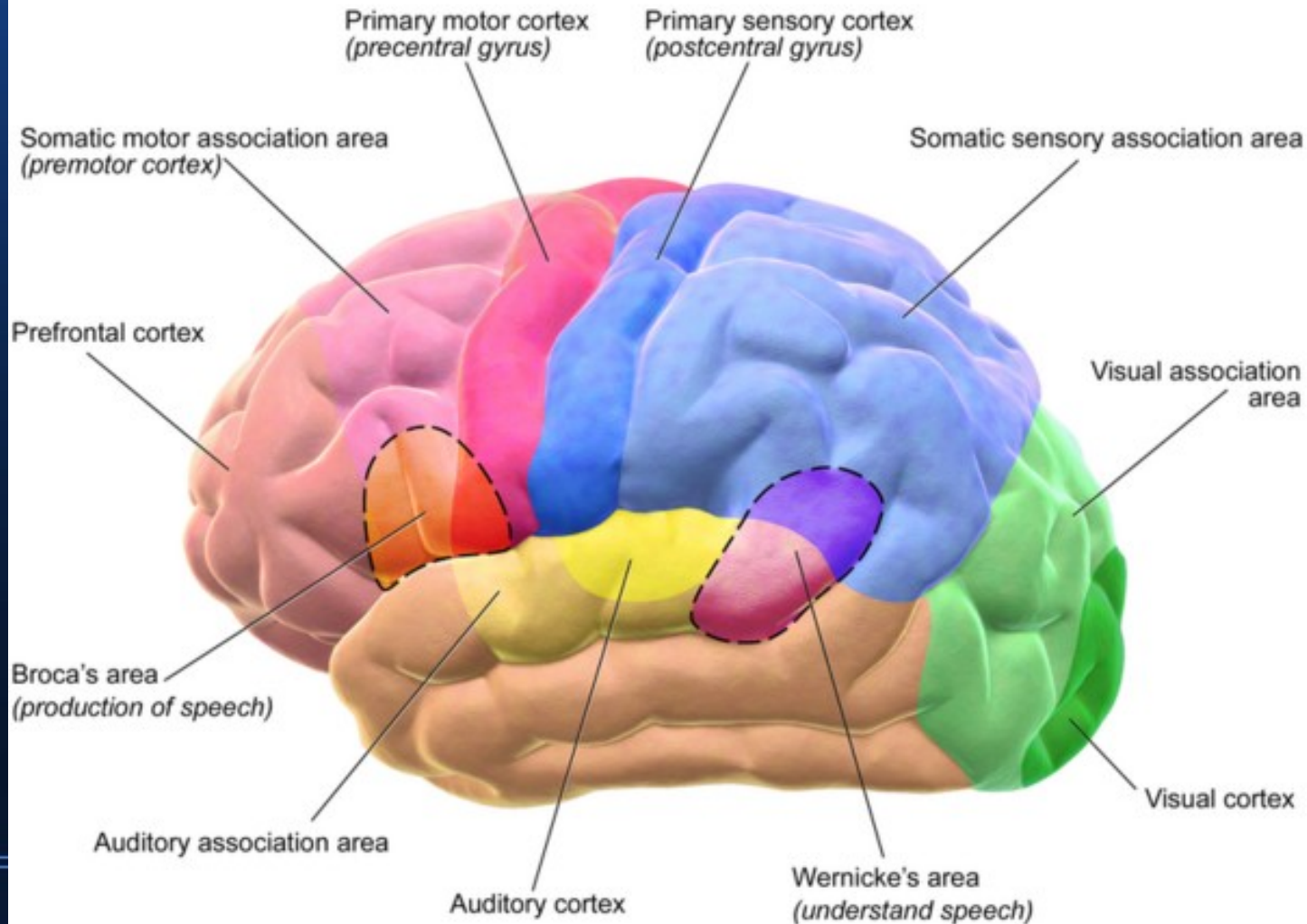
後來 Edgar Adrian 又量測皮膚的電衝動

- 結果發現皮膚在穩定的刺激下，會產生強弱相同的訊號，但是訊號的頻率卻會隨著時間而降低。腦部的感應也會隨之降低。

A key result, published in 1928, stated that the excitation of the skin under constant stimulus is initially strong but gradually decreases over time, whereas the sensory impulses passing along the nerves from the point of contact are constant in strength, yet are reduced in frequency over time, and the sensation in the brain diminishes as a result.

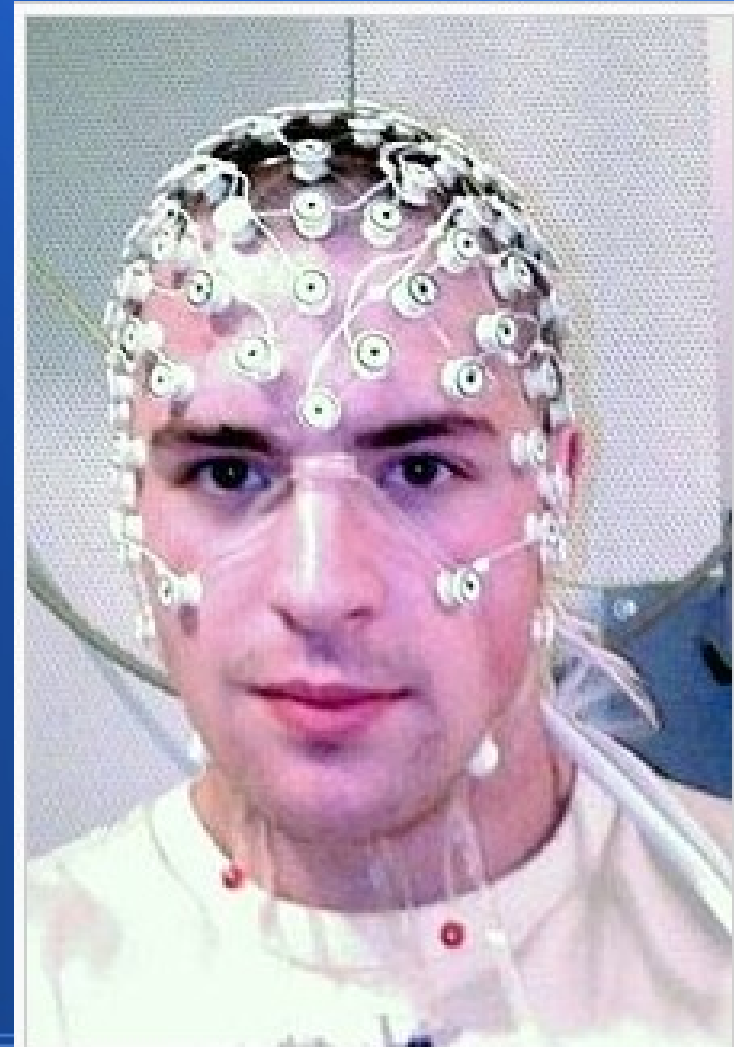
接著他繼續研究而繪製出了
一些動物大腦皮質的感覺地圖

Motor and Sensory Regions of the Cerebral Cortex



最後 Edgar Adrian 又用腦電圖 (EEG)

- 研究人類的腦訊號，並找出了癲癇和嗅覺等訊號的特性。



腦電圖記錄的設置



抱歉、話題扯遠了

- 讓我們拉回到神經網路的主題。
- 將焦點放在神經元間連接點的突觸上面！

這些突觸之間

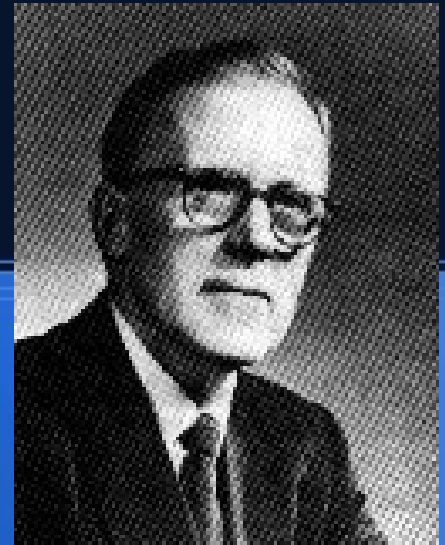
- 會因為長期高頻率活動而得以增加，這一效應稱為長期增強作用。
- 一種稱為 NMDA 的受體通道被認為與長期增強作用密切相關。

Ramón Cajal 認為

- 學習機制並不需要形成新神經元。
- 他於 1894 年在倫敦皇家內科醫學院發表的演講中，他提出，記憶可能是由加強現有神經元之間的聯繫、從而提高它們溝通的有效性而形成的

上述這個想法

- 啟發了 Donald Olding Hebb
- Hebb 在 1949 年在《行為的組織》（ The Organization of Behavior ）一書中，提出《細胞可能通過構建新的連接或經歷代謝變化而提高它們的溝通能力》
- 這個理論被稱為《赫布理論》（Hebbian theory）



唐納德·赫布

出生	1904年7月22日 加拿大新斯科舍省切斯特
逝世	1985年8月20日（81歲） 加拿大新斯科舍省切斯特
居住地	加拿大
國籍	 加拿大
研究領域	心理學
任職於	蒙特婁神經科學研究所 女王大學 耶基斯國家靈長類研究中心 麥吉爾大學 ^[1]
母校	戴爾豪斯大學 麥吉爾大學 哈佛大學
著名成就	細胞結集理論
獲獎	皇家學會院士 ^[2]

更詳細的赫布理論如下

赫布理論（英語：**Hebbian theory**）是一個**神經科學**理論，解釋了在**學習**的過程中**腦**中的**神經元**所發生的變化。赫布理論描述了**突觸可塑性**的基本原理，即突觸前神經元向突觸後神經元的持續重複的刺激，可以導致突觸傳遞效能的增加。這一理論由**唐納德·赫布**於1949年提出^[1]，又被稱為**赫布定律**（Hebb's rule）、**赫布假說**（Hebb's postulate）、**細胞結集理論**（cell assembly theory）等。他如此表述這一理論：

我們可以假定，反射活動的持續與重複會導致**神經元**穩定性的持久性提升……當神經元A的**軸突**與神經元B很近並參與了對B的重複持續的興奮時，這兩個神經元或其中一個便會發生某些生長過程或代謝變化，致使A作為能使B興奮的細胞之一，它的效能增強了。

這一理論經常會被總結為「一起發射的神經元連在一起」（*Cells that fire together, wire together*）^[2]。但是，這個解釋並不完整，赫布強調說，神經元"A"必須「對神經元"B"的發射「作出了一定貢獻」，因此，神經元"A"的發射必須在神經元"B"之先，而不能同時發射。赫布理論中的這一部分研究，後來被稱作**STDP**，表明突觸可塑性需要一定的時間延遲。^[3]赫布理論可以用於解釋「聯合學習」（**associative learning**），在這種學習中，由對神經元的重複刺激，使得神經元之間的突觸強度增加。這樣的學習方法被稱為**赫布型學習**（Hebbian learning）。赫布理論也成為了**無誤性學習方法**的生物學基礎。

後來這個理論繼續被延伸

赫布記憶印痕與細胞結集理論 [編輯]

赫布理論解釋了神經元如何組成聯接，從而形成**記憶印痕**。赫布理論闡明了細胞集群（Cell Assemblies）的形態和功能：「兩個神經元或者神經元系統，如果總是同時興奮，就會形成一種『組合』，其中一個神經元的興奮會促進另一個的興奮。」^{[1]:70} 赫布同時寫道：「如果一個神經元持續激活另一個神經元，前者的軸突將會生長出突觸小體（如果已有，則會繼續長大）和後者的胞體相連接。」^{[1]:63}

高爾頓·威拉德·奧爾波特根據自聯想(auto-association)的思路，進一步提出了關於細胞集群的作用，以及它們在形成記憶痕跡中的角色：

如果系統的輸入會導致同樣的模式重複出現，那麼組成這個模式的元素之間的相互關聯性將會大大增強。這意味著，其中任何一個元素都會傾向於觸發同組的其他元素，同時（以減少權重的方式）抑制組外其他不相關元素。另一個角度來看，這個模式作為一個整體實現了「自聯想」。我們可以把一個學習了（自聯想）的模式稱為記憶痕跡。^{[4]:44}

通常認為，從整體的角度來看，赫布學習是**神經網絡**形成記憶痕跡的首要基礎。

埃里克·坎德爾在一種稱作**加州海兔**的海生**腹足綱**動物上關於突觸學習機制的研究，進一步表明了赫布理論和這一學習機制的關聯。

在**脊椎動物**的**中樞神經系統**中進行關於赫布突觸可塑機制的實驗，要比在海洋無脊椎動物的**周邊神經系統**中難以控制許多。許多關於脊椎動物神經元的長時期突觸變化（例如**長期增強作用**）的實驗，都是對腦細胞的模擬而非進行真實的生理學實驗。儘管如此，一些關於脊椎動物的腦內突觸可塑機制的生理實驗，也傾向於符合赫布理論。

接著被用到《人工智慧》領域

公式 [編輯]

在人工神經網絡中，突觸間傳遞作用的變化被當作是（被映射成）神經元網絡圖中相應權重的變化。如果兩個神經元同步激發，則它們之間的權重增加；如果單獨激發，則權重減少。赫布學習規則是最古老的也是最簡單的神經元學習規則。

以下是一種描述赫布學習的公式：

$$w_{ij} = x_i x_j$$

其中 w_{ij} 是從神經元 j 到神經元 i 的聯接權重， x_i 是神經元 i 的輸入。注意這是模式學習，每個訓練樣本都會導致權重改變。在Hopfield神經網絡中，如果 $i = j$ 則 w_{ij} 恆為0（沒有神經元和自身相連）。對於二進位的神經元（激發值只能為0或者1），如果它聯接的神經元有相同的激發模式，則聯接會設定為1。

另一種表達式為：

$$w_{ij} = \frac{1}{p} \sum_{k=1}^p x_i^k x_j^k,$$

其中 w_{ij} 是從神經元 j 到神經元 i 的聯接權重， p 是訓練模式的個數，而 x_i^k 是神經元 i 的第 k 個輸入。（這是按照時段來進行學習，在所有訓練樣本都給出了之後再更新權重。）同樣的，在Hopfield神經網絡中，如果 $i = j$ 則 w_{ij} 恆為0（沒有神經元和自身相連）。

赫布理論有一個變體，稱作哈里·克洛普弗理論（Harry Klopf's model），考慮了包括阻斷等生理現象，可以再現很多生物學表現，也很容易使用。

這就是《赫布學習法則》

在人工神經網絡中，突觸間傳遞作用的變化被當作是（被映射成）神經元網絡圖中相應權重的變化。如果兩個神經元同步激發，則它們之間的權重增加；如果單獨激發，則權重減少。赫布學習規則是最古老的也是最簡單的神經元學習規則。

以下是一種描述赫布學習的公式：

$$w_{ij} = x_i x_j$$

其中 w_{ij} 是從神經元 j 到神經元 i 的聯接權重， x_i 是神經元 i 的輸入。注意這是模式學習，每個訓練樣本都會導致權重改變。在Hopfield神經網絡中，如果 $i = j$ 則 w_{ij} 恆為0（沒有神經元和自身相連）。對於二進位的神經元（激發值只能為0或者1），如果它聯接的神經元有相同的激發模式，則聯接會設定為1。

Hebb's Rule

另一種表達式為：

$$w_{ij} = \frac{1}{p} \sum_{k=1}^p x_i^k x_j^k$$

其中 w_{ij} 是從神經元 j 到神經元 i 的聯接權重， p 是訓練模式的個數，而 x_i^k 是神經元 i 的第 k 個輸入。（這是按照時段來進行學習，在所有訓練樣本都給出了之後再更新權重。）同樣的，在Hopfield神經網絡中，如果 $i = j$ 則 w_{ij} 恆為0（沒有神經元和自身相連）。

赫布學習法則

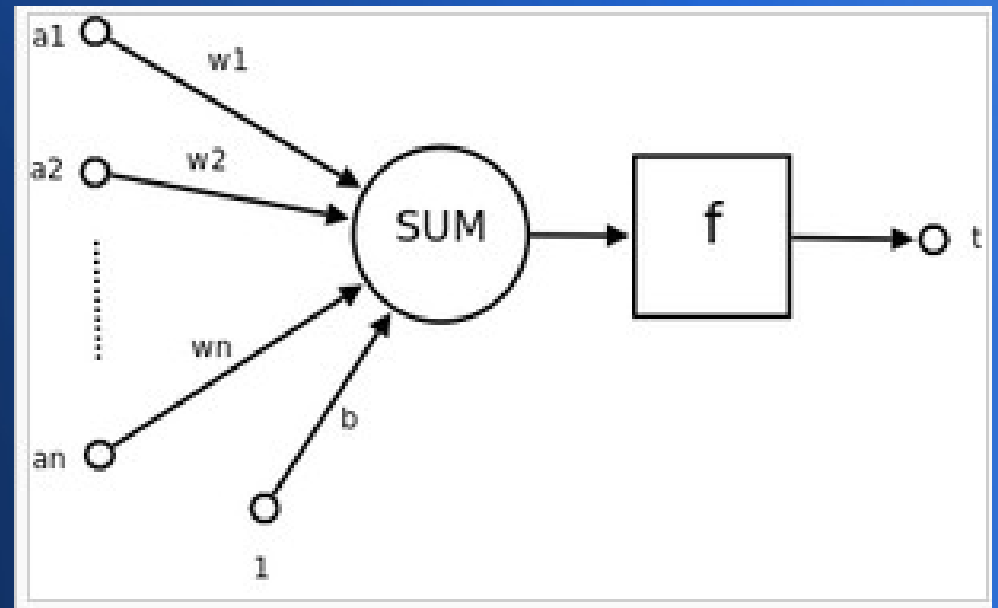
- 對人工智慧先驅們有很大的啟發

於是

- 有些神經網路程式被發展出來了

首先是

- Frank Rosenblatt 在 1957 年在於 Cornell 航空實驗室發明了《感知器》（Perceptron）
- 但是他們只能透過《赫布法則》最小化《單層網路》的權重。

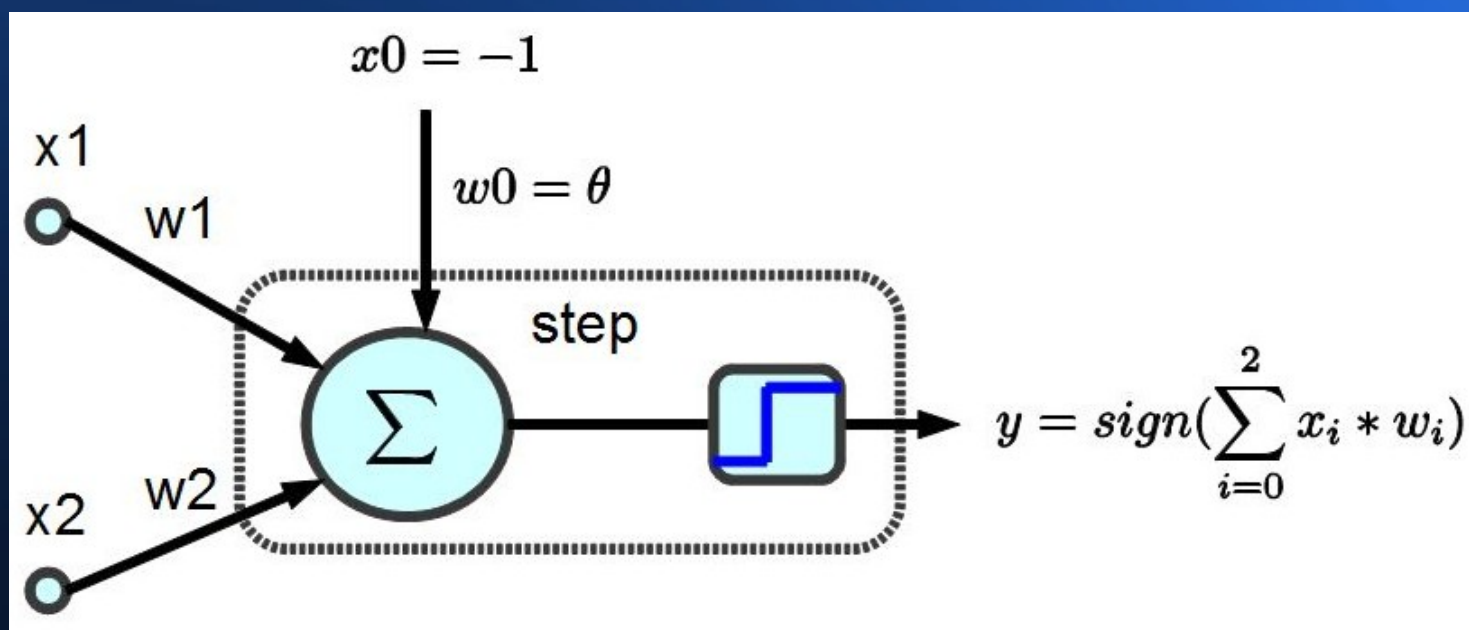


$$t = f\left(\sum_{i=1}^n w_i x_i + b\right) = f(\mathbf{w}^T \mathbf{x})$$

$$f(n) = \begin{cases} +1 & \text{if } n \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

像是以下圖形

- 是《兩輸入的單層感知器》的結構



程式實作：<http://ccc.nqu.edu.tw/wd.html#ai/perceptron.wd>

當我們用感知器

- 學習一些基本的邏輯閘函數時

```
var andTable = [ [ 0, 0, 0 ], [ 0, 1, 0 ], [ 1, 0, 0 ], [ 1, 1, 1 ] ]; // AND 函數的真值表
var orTable  = [ [ 0, 0, 0 ], [ 0, 1, 1 ], [ 1, 0, 1 ], [ 1, 1, 1 ] ]; // OR 函數的真值表
var xorTable = [ [ 0, 0, 0 ], [ 0, 1, 1 ], [ 1, 0, 1 ], [ 1, 1, 0 ] ]; // XOR 函數的真值表

learn("and", andTable); // 學習 AND 函數
learn("or",  orTable);  // 學習 OR 函數
learn("xor", xorTable); // 學習 XOR 函數
```


會發現有些成功有些失敗

● AND :

```
34:x=(-1.000, 0.000, 0.000) w=(0.660, 0.340, 0.340) y=0.000 yd=0.000 e=0.000
34:x=(-1.000, 0.000, 1.000) w=(0.660, 0.340, 0.340) y=0.000 yd=0.000 e=0.000
34:x=(-1.000, 1.000, 0.000) w=(0.660, 0.340, 0.340) y=0.000 yd=0.000 e=0.000
34:x=(-1.000, 1.000, 1.000) w=(0.660, 0.340, 0.340) y=1.000 yd=1.000 e=0.000
學習成功 !
```

● OR :

```
24:x=(-1.000, 0.000, 0.000) w=(0.380, 0.380, 0.380) y=0.000 yd=0.000 e=0.000
24:x=(-1.000, 0.000, 1.000) w=(0.380, 0.380, 0.380) y=1.000 yd=1.000 e=0.000
24:x=(-1.000, 1.000, 0.000) w=(0.380, 0.380, 0.380) y=1.000 yd=1.000 e=0.000
24:x=(-1.000, 1.000, 1.000) w=(0.380, 0.380, 0.380) y=1.000 yd=1.000 e=0.000
學習成功 !
```

● XOR :

```
99:x=(-1.000, 0.000, 0.000) w=(0.090, 0.080, 0.080) y=0.000 yd=0.000 e=0.000
99:x=(-1.000, 0.000, 1.000) w=(0.080, 0.080, 0.090) y=0.000 yd=1.000 e=1.000
99:x=(-1.000, 1.000, 0.000) w=(0.080, 0.080, 0.090) y=1.000 yd=1.000 e=0.000
99:x=(-1.000, 1.000, 1.000) w=(0.090, 0.070, 0.080) y=1.000 yd=0.000 e=-1.000
學習失敗 !
```

學者們發現

- 《單層感知器》雖然可以做出一些簡單的電路，像是 AND, OR, NOT
- 像是 Minsky 就證明了《單層感知器》不可能學會像 XOR 這麼簡單的函數。

所以

- 學術界對神經網路的激情就慢慢的消退了！

但是仍然有人繼續發展

例如在 1982 年 Hopfield 提出的《霍普菲爾德》神經網路

- 幾乎就是完全按照《赫布理論》運作的

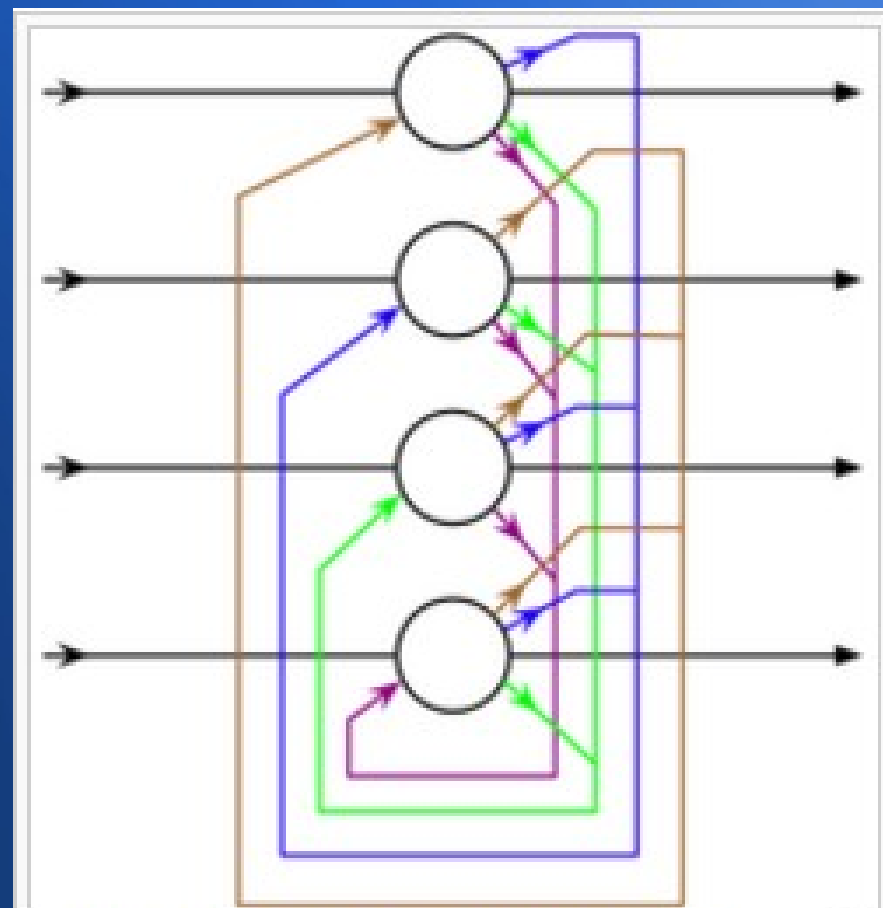
Hopfield 網絡的连接有以下特徵：

- $w_{ii} = 0, \forall i$ (沒有神經元和自身相連)
- $w_{ij} = w_{ji}, \forall i, j$ (連接權重是對稱的)

$$s_i \leftarrow \begin{cases} +1 & \text{if } \sum_j w_{ij}s_j \geq \theta_i, \\ -1 & \text{otherwise.} \end{cases}$$

where:

- w_{ij} is the strength of the connection weight from unit j to unit i
- s_j is the state of unit j .
- θ_i is the threshold of unit i .



一個有四個節點的Hopfield網絡。

Hopfield 神經網路所採用的學習法則如下

Hebbian learning rule for Hopfield networks [\[edit \]](#)

The [Hebbian Theory](#) was introduced by Donald Hebb in 1949, in order to explain "associative learning", in which simultaneous activation of neuron cells leads to pronounced increases in synaptic strength between those cells.^[4] It is often summarized as "Neurons that fire together, wire together. Neurons that fire out of sync, fail to link".

The Hebbian rule is both local and incremental. For the Hopfield Networks, it is implemented in the following manner, when learning n binary patterns:

$$w_{ij} = \frac{1}{n} \sum_{\mu=1}^n \epsilon_i^{\mu} \epsilon_j^{\mu}$$

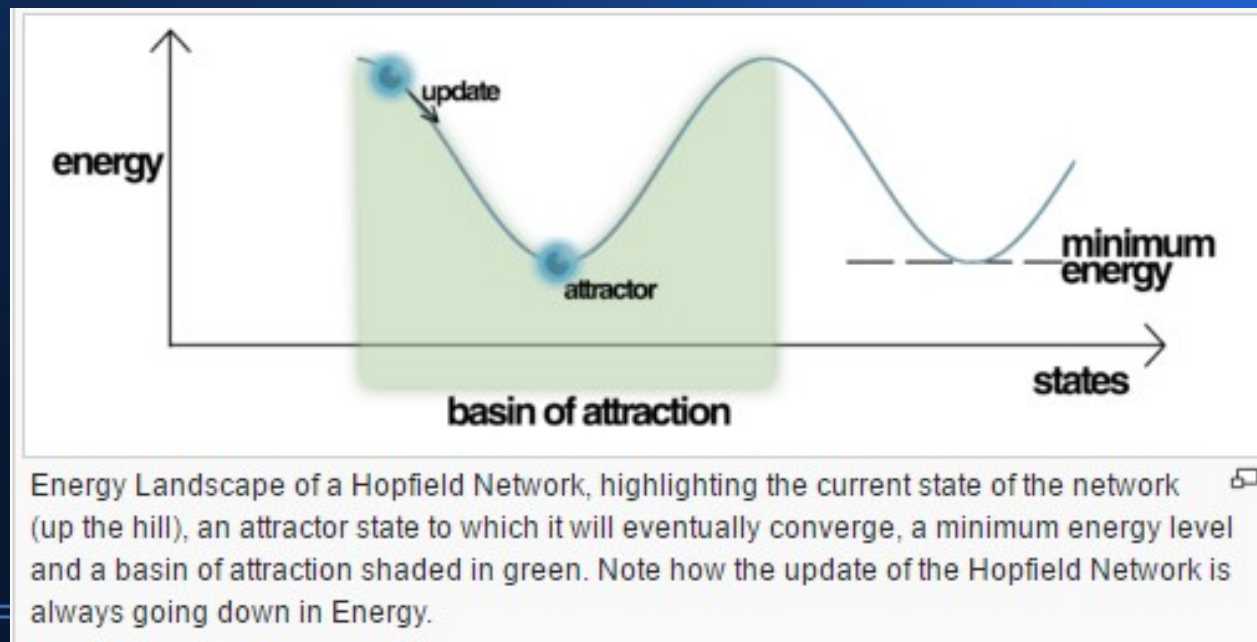
where ϵ_i^{μ} represents bit i from pattern μ .

If the bits corresponding to neurons i and j are equal in pattern μ , then the product $\epsilon_i^{\mu} \epsilon_j^{\mu}$ will be positive. This would, in turn, have a positive effect on the weight w_{ij} and the values of i and j will tend to become equal. The opposite happens if the bits corresponding to neurons i and j are different.

根據數學推導

- Hopfield 神經網路最後會落在下列能量函數的最低點

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j + \sum_i \theta_i s_i$$



所以 Hopfield 神經網路

- 仍然是個數學上的優化問題
- 只不過是透過《赫布法則》將
能量函數最小化而已！

仔細看看這兩個數學式

- Hopfield 網路能量函數

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j + \sum_i \theta_i s_i$$

- Hebb's Rule (赫布學習法則)

$$w_{ij} = \frac{1}{n} \sum_{\mu=1}^n \epsilon_i^{\mu} \epsilon_j^{\mu}$$

where ϵ_i^{μ} represents bit i from pattern μ .

您會發現霍普菲爾德網路

- 透過《赫布學習法則》來最小化其能量式

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j + \sum_i \theta_i s_i$$

- 這種網路具有《容錯的記憶性》，所以當您的輸入有些位元錯誤了，還是可以被網路復原回正確的結果。

但是霍普菲爾德網路

- 比較像是《容錯記憶元件》
- 不過《容錯記憶》和《智慧推理》兩者，
其實可說是一體的兩面！

後來在 1986 年

- David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams 等人用《反向傳播學習法》(Backpropagation) 做出了很好的結果。才讓《多層感知器》開始廣為人知，於是神經網路領域又變得炙手可熱。

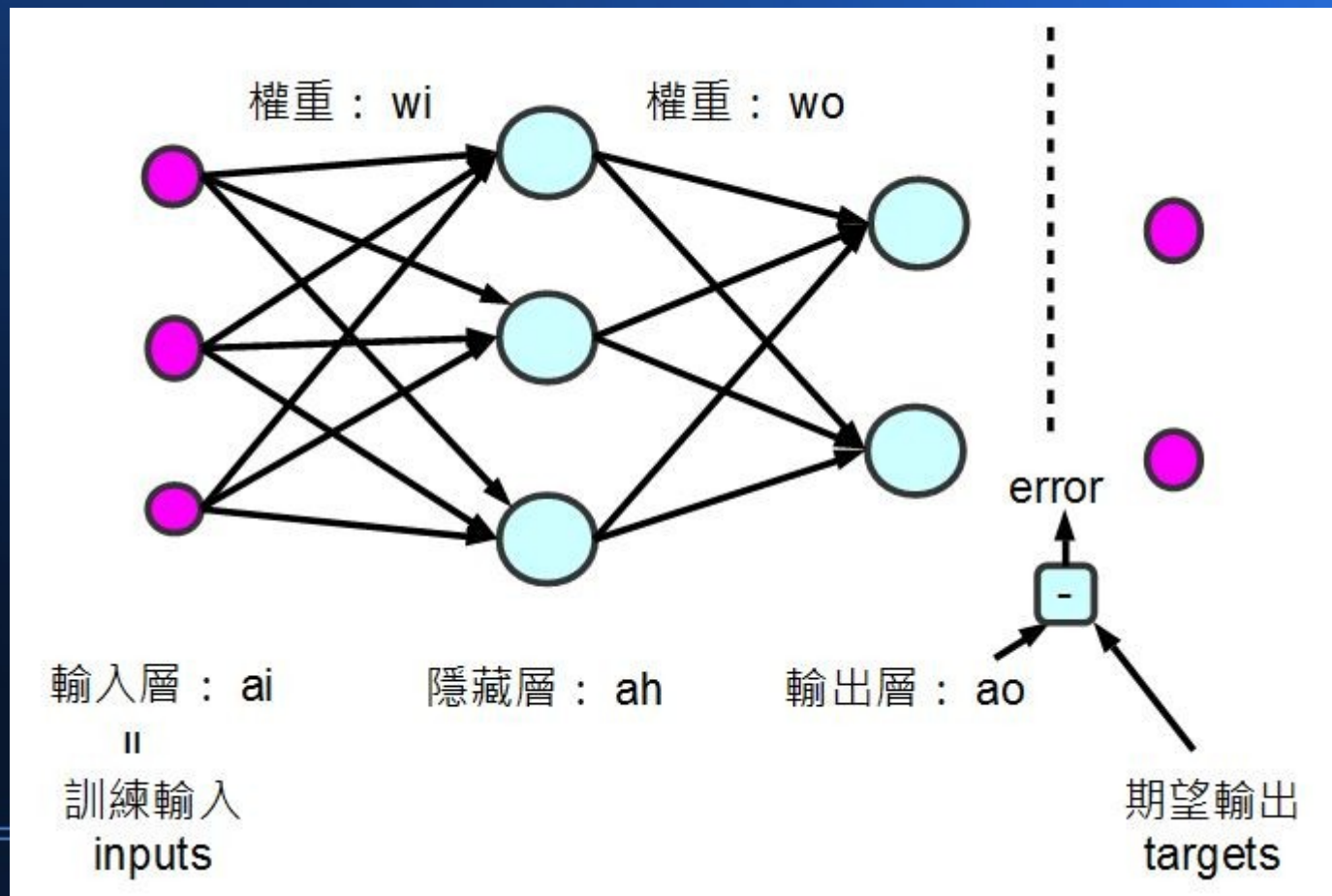
Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (8 October 1986). "Learning representations by back-propagating errors". *Nature*. 323 (6088): 533–536.

那個重要的《反向傳播學習法》

- 一直到現在都還非常有用
- 而且是《深度學習網路》的兩大重要算法之一。

反傳遞演算法

- 讓多層感知器 MLP 能有效的被訓練

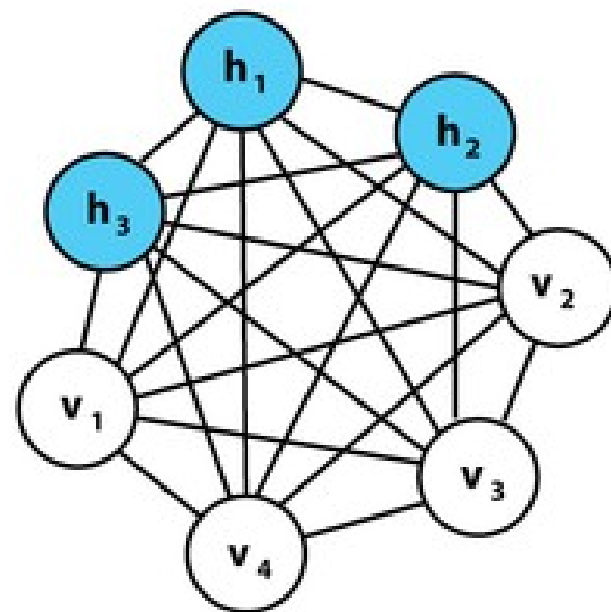


事實上

- 反向傳播演算法並非 Hinton 等人所創造
- 而是在 1960 年代就由控制領域的 Henry J. Kelley 最早提出來，後來 Arthur E. Bryson, Stuart Dreyfus , Vapnik 等人都有些貢獻
- 1982 年 Paul Werbos 把這個方法用在神經網路模型上。
然後才是 1986 年 Hinton 等人創造出很好的應用案例。

Hinton 的貢獻不只如此

- 他和 Terry Sejnowski 在 1985 發明了一種稱為《波茲曼機 Boltzmann Machine》的神經網路，



A graphical representation of an example Boltzmann machine. Each undirected edge represents dependency. In this example there are 3 hidden units and 4 visible units. This is not a restricted Boltzmann machine.

這種波茲曼機

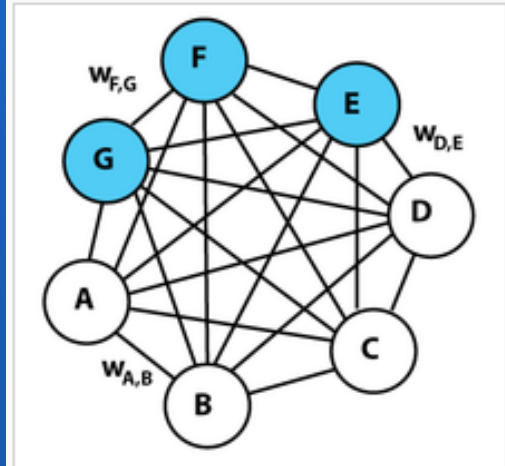
- 和《霍普菲爾德網路》很像，但卻引入了《機率模型》

A Boltzmann machine, like a Hopfield network, is a network of units with an "energy" defined for the network. It also has binary units, but unlike Hopfield nets, Boltzmann machine units are stochastic. The global energy, E , in a Boltzmann machine is identical in form to that of a Hopfield network:

$$E = - \left(\sum_{i,j} w_{ij} s_i s_j + \sum_i \theta_i s_i \right)$$

Where:

- w_{ij} is the connection strength between unit j and unit i .
- s_i is the state, $s_i \in \{0, 1\}$, of unit i .
- θ_i is the bias of unit i in the global energy function. ($-\theta_i$ is the activation threshold for the unit.)



A graphical representation of a Boltzmann machine with a few weights labeled. Each undirected edge represents dependency and is weighted with weight w_{ij} . In this example there are 3 hidden units (blue) and 4 visible units (white). This is not a restricted Boltzmann machine.

這種引入機率模型的神經網路

- 稱為《信念網路》
(Belief Network)

不過波茲曼機表現不算太好

- 學習演算法的速度很慢！

$$\Delta E_i = \sum_j w_{ij} s_j + \theta_i$$

$$\Delta E_i = -k_B T \ln(p_{i=\text{off}}) - (-k_B T \ln(p_{i=\text{on}}))$$

$$p_{i=\text{on}} = \frac{1}{1 + \exp(-\frac{\Delta E_i}{T})}$$

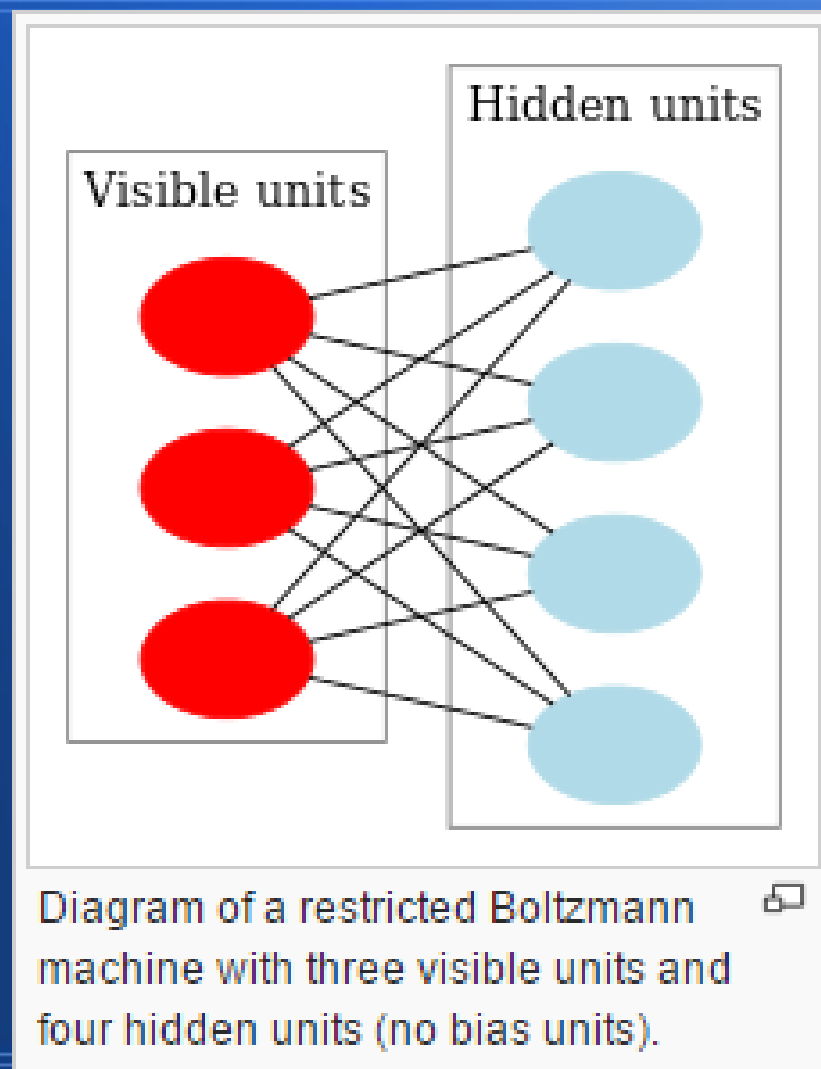
$$\begin{aligned}\frac{\Delta E_i}{T} &= \ln(p_{i=\text{on}}) - \ln(p_{i=\text{off}}) \\ \frac{\Delta E_i}{T} &= \ln(p_{i=\text{on}}) - \ln(1 - p_{i=\text{on}}) \\ \frac{\Delta E_i}{T} &= \ln\left(\frac{p_{i=\text{on}}}{1 - p_{i=\text{on}}}\right) \\ -\frac{\Delta E_i}{T} &= \ln\left(\frac{1 - p_{i=\text{on}}}{p_{i=\text{on}}}\right) \\ -\frac{\Delta E_i}{T} &= \ln\left(\frac{1}{p_{i=\text{on}}} - 1\right) \\ \exp\left(-\frac{\Delta E_i}{T}\right) &= \frac{1}{p_{i=\text{on}}} - 1\end{aligned}$$

所以當時

- 沒有得到很多研究者青睞！

但是在 1986 年

- Paul Smolensky 改進《波茲曼機》為《受限波茲曼機》(Restricted Boltzmann Machine)，分為顯層和隱層
- 連結只出現在兩層之間，同一層的節點之間不會有連結



波茲曼機的發明人 Hinton

- 後來注意到《受限波茲曼機》(RBM) 的模型，並且於 2002 年創造出一個快速的 RBM 學習算法，稱為《對比分歧法》
(Contrast Divergence, CD)

對比分歧 CD 學習算法如下

受限玻爾茲曼機的訓練目標是針對某一訓練集 V ，最大化機率的乘積。其中， V 被視為一矩陣，每個行向量作為一個可見單元向量 v ：

$$\arg \max_W \prod_{v \in V} P(v)$$

或者，等價地，最大化 V 的對數機率期望：^{[9][10]}

$$\arg \max_W \mathbb{E} \left[\sum_{v \in V} \log P(v) \right]$$

訓練受限玻爾茲曼機，即最優化權重矩陣 W ，最常用的算法是傑弗里·辛頓提出的對比分歧（contrastive divergence，CD）算法。這一算法最早被用於訓練辛頓提出的「專家積」模型^[12]。這一算法在梯度下降的過程中使用吉布斯採樣完成對權重的更新，與訓練前饋神經網絡中利用反向傳播算法類似。

基本的針對一個樣本的單步對比分歧（CD-1）步驟可被總結如下：

1. 取一個訓練樣本 v ，計算隱層節點的機率，在此基礎上從這一機率分布中獲取一個隱層節點激活向量的樣本；
2. 計算 v 和 h 的外積，稱為「正梯度」；
3. 從 h 獲取一個重構的可見層節點的激活向量樣本 v' ，此後從 v' 再次獲得一個隱層節點的激活向量樣本 h' ；
4. 計算 v' 和 h' 的外積，稱為「負梯度」；
5. 使用正梯度和負梯度的差以一定的學習率更新權重 $w_{i,j}$ ： $\Delta w_{i,j} = \epsilon(vh^T - v'h'^T)$ 。

偏置 a 和 b 也可以使用類似的方法更新。

結果讓《受限波茲曼機》

- 在《維度降低、分類、過濾、特徵抽取》等領域有了很好的結果。

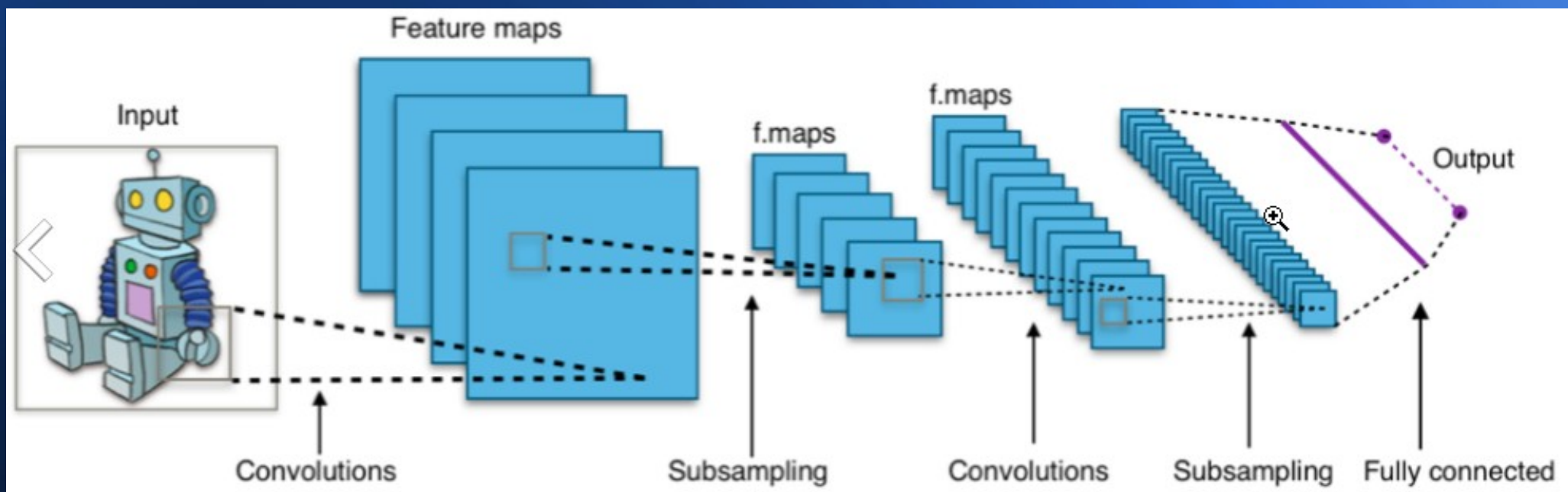
更厲害的是

- 《受限波茲曼機》多層套疊仍然可以運作良好。
- 還可以和《多層感知器》有效疊合，形成《深度神經網路》。

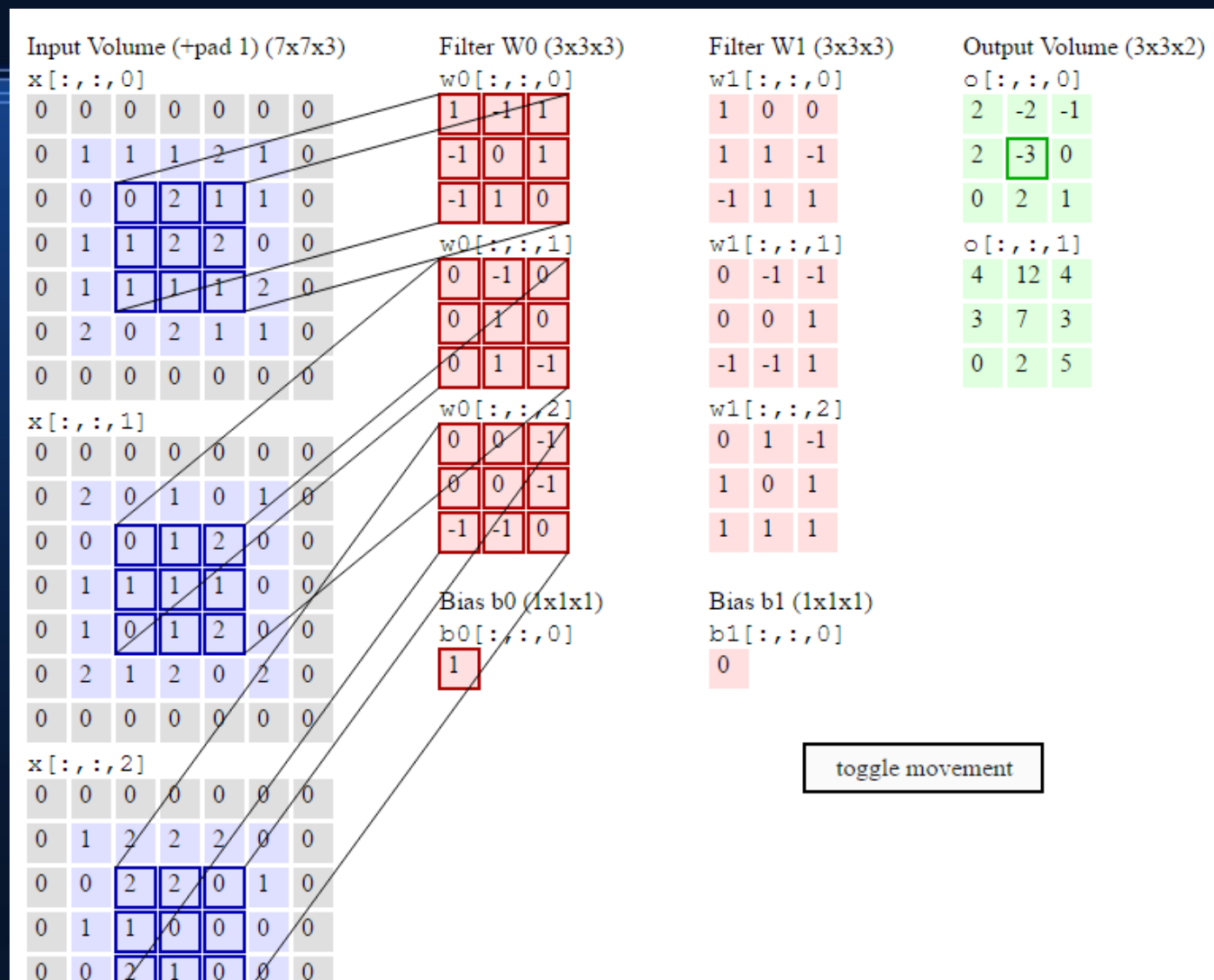
後來在影像識別領域

- 這種《深度神經網路》被加入了《捲積 Convolutional 》運算
- 成為《卷積神經網絡 Convolutional Neural Network, CNN 》

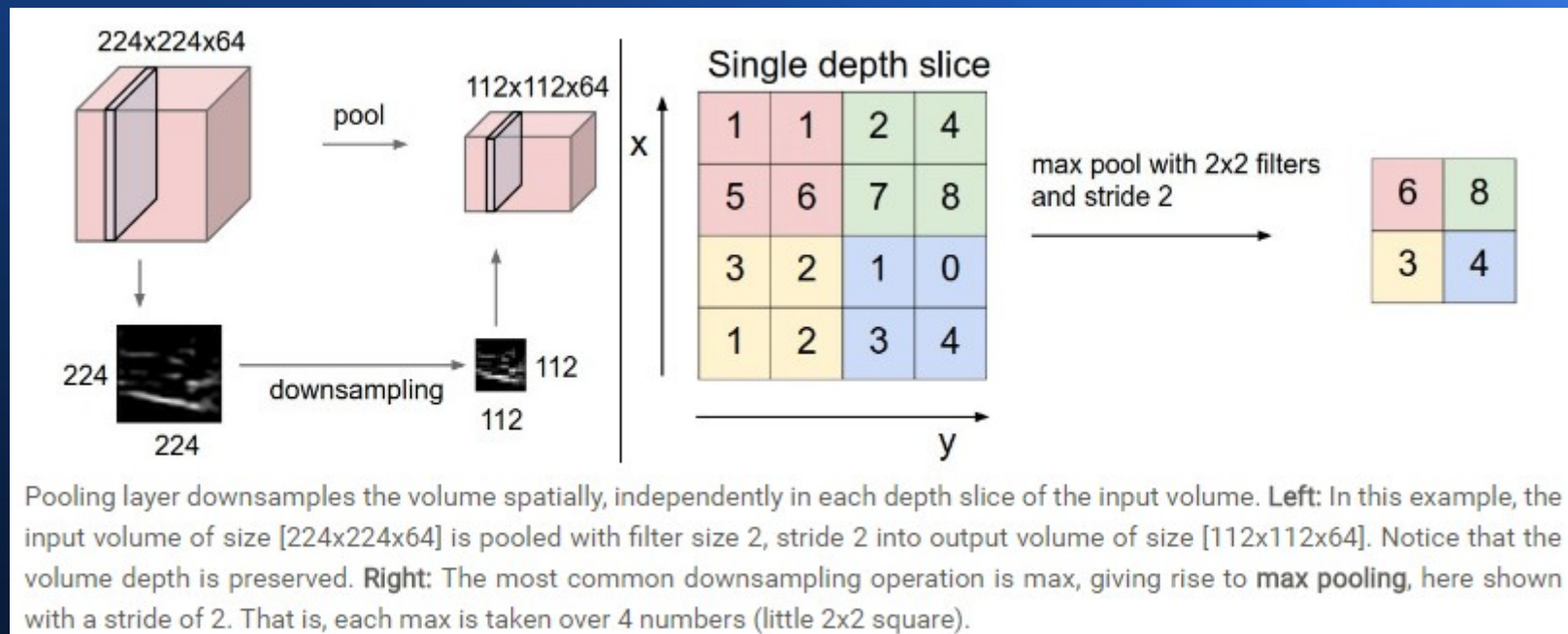
捲積神經網路的架構如下



捲積時會一塊一塊進行遮罩加總



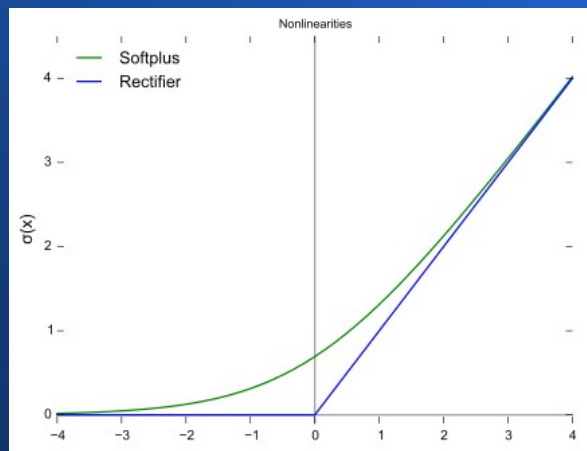
然後進行池化降維運算



另外還會加入 RELU 函數連接

線性整流函數 (Rectified Linear Unit, **ReLU**), 又稱修正線性單元, 是一種人工神經網絡中常用的激活函數 (activation function), 通常指代以斜坡函數及其變種為代表的非線性函數。

比較常用的線性整流函數有斜坡函數 $f(x) = \max(0, x)$, 以及帶泄露整流函數 (Leaky ReLU), 其中 x 為神經元 (Neuron) 的輸入。線性整流被認為有一定的生物學原理^[1], 並且由於在實踐中通常有著比其他常用激活函數 (譬如邏輯函數) 更好的效果, 而被如今的深度神經網絡廣泛使用於諸如圖像識別等計算機視覺^[1]人工智慧領域。

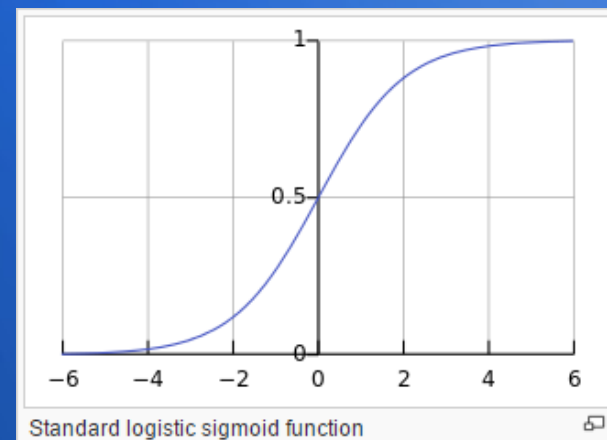


直線版 Rectifier

$$f(x) = \max(0, x)$$

平滑版 Softplus

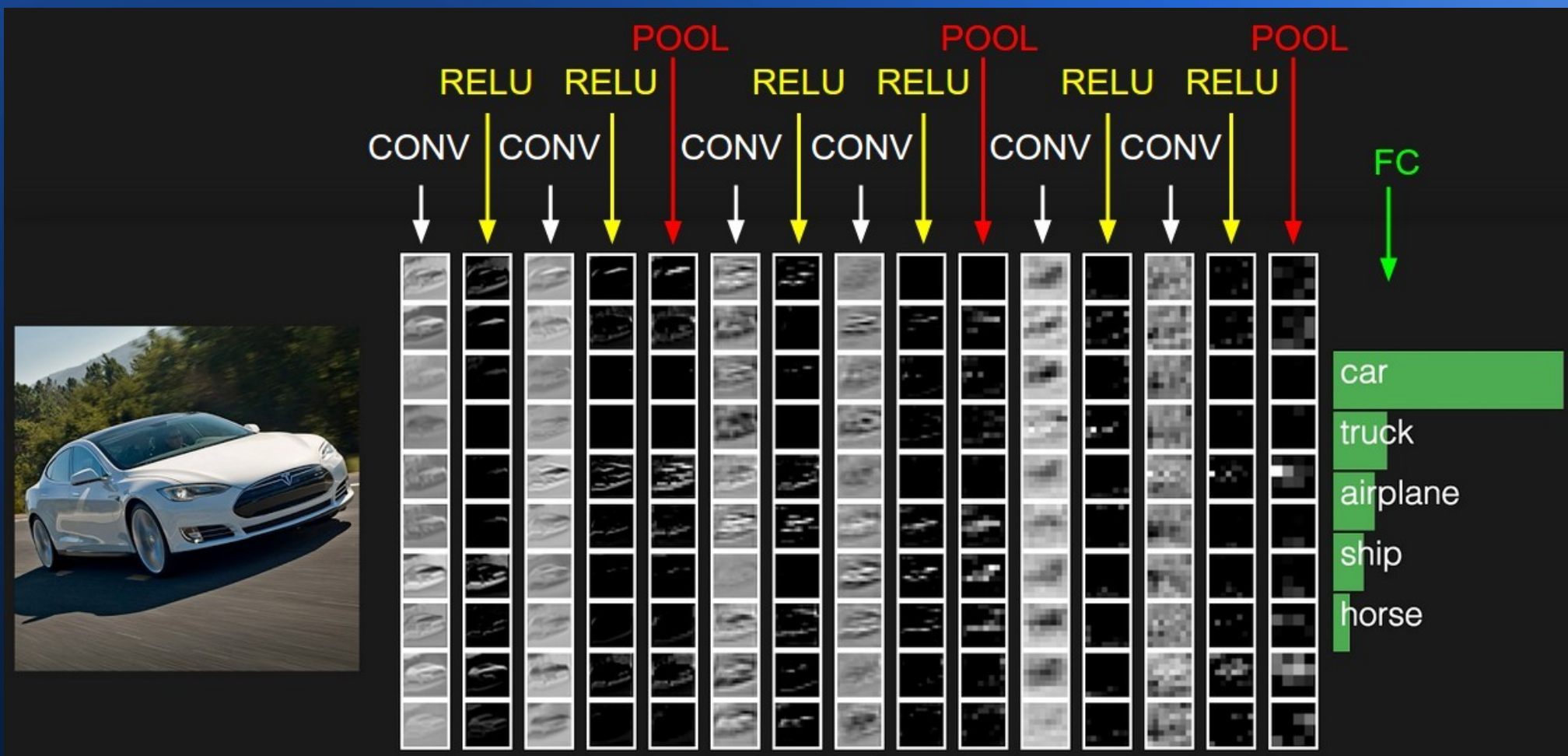
$$f(x) = \ln(1 + e^x)$$



註：這和傳統神經網路使用 Sigmoid 的《邏輯 logistic》函數有所不同

$$f(x) = \frac{1}{1 + e^{-x}}$$

於是形成影像捲積神經網路



這三種技術疊合的 《深度神經網路技術》

- 造成了整個《人工智慧領域》重大的
進展！

這些重大進展

- 已經對《影像處理、電腦圍棋、機器翻譯領域》造成很大改變
- 未來還不知道會往那個方向前進！

因而造成了最近幾年

- 所謂《深度學習 Deep Learning》
技術的興起！

我們將在下次的十分鐘系列裏

- 介紹深度學習技術的方法！

希望你會喜歡

今天的十分鐘系列

我們下次見！

Bye Bye!