

說得越短，懂得越多

本課講 1993 年的一篇論文。作者是 Hinton 和 van Camp。他們讓電腦的每個「旋鈕」只記個大概，電腦反而學得更好。

適合 10 歲左右

約 40 分鐘

需要：鉛筆、白紙、一位同伴

1 熱身：用最少的字說清楚

下面每一行都是一串圖形。請你用最少的字，把它說清楚。別人聽了你的話，要能一個不差地寫出來。

- ① ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ _____
- ② ○ △ ○ △ ○ △ ○ △ ○ △ _____
- ③ ○ ○ ○ ○ ○ ○ △ ○ ○ ○ _____
- ④ △ ○ ○ △ △ △ ○ △ ○ ○ _____

① 很好說：「10 個圓」。② 也行：「圓三角，重複 5 次」。③ 要多說一點：「10 個圓，第 7 個換成三角」。④ 呢？你找不到規律，只能一個一個念。

發現了嗎？**有規律的東西，可以說得很短**。沒有規律的東西，只能一個一個死記。

2 好的學習 = 用最短的話說清楚

老師想知道：今天班上 30 個同學，誰帶了水杯？班長可以用 3 種方法報告。

方法	班長怎麼說	說了多長
A. 全背下來	「張一帶了，李二帶了……」念 30 個名字	很長
B. 規則太粗	「沒人帶。除了張一，李二……」念 28 個名字	還是很長
C. 好規則	「都帶了，只有小明和小紅沒帶。」	很短

方法 C = 一條短規則（都帶了）+ 幾個例外（小明、小紅）。規則越短，例外越少，加起來就越短。加起來最短的說法，往往就是**最懂的說法**。

好規則還能猜新情況。明天來了一個轉學生，方法 C 的班長會猜：「他大概也帶了。」方法 A 的班長只背過 30 個名字，沒法猜。

說清楚一件事要用多少字，科學家叫它**描述長度**。這篇論文的想法是：電腦學東西時，也要讓描述長度越短越好。

3 死記硬背的電腦

你見過只背答案的同學嗎？他背下「第 1 題選 B，第 2 題選 C」。練習冊上的題，他全做對了。可是考試換了新題，他就不會了。

電腦也會犯這個毛病。它把練習題的答案全背下來，連題目裡碰巧出現的小細節也背下來。練習時它滿分，遇到新題卻錯一大片。這個毛病叫**過擬合**。你可以把它記成「死記硬背」。

能記的東西多，練習題少，最容易死記硬背。怎麼防止？Hinton 的答案是：**別讓它記得太精確**。要明白這句話，先認識電腦裡的「旋鈕」。

4 電腦裡的旋鈕

第 1 課說過：電腦猜錯了，就「悄悄調整自己」。它調的是什麼？是成千上萬個**旋鈕**。

旋鈕就像收音機上調音量的按鈕。每個旋鈕擰到一個位置，就是一個數。所有旋鈕一起決定電腦怎麼回答問題。科學家管這些旋鈕叫**權重**。



精確：3.1415926... 模糊：大概 3 左右

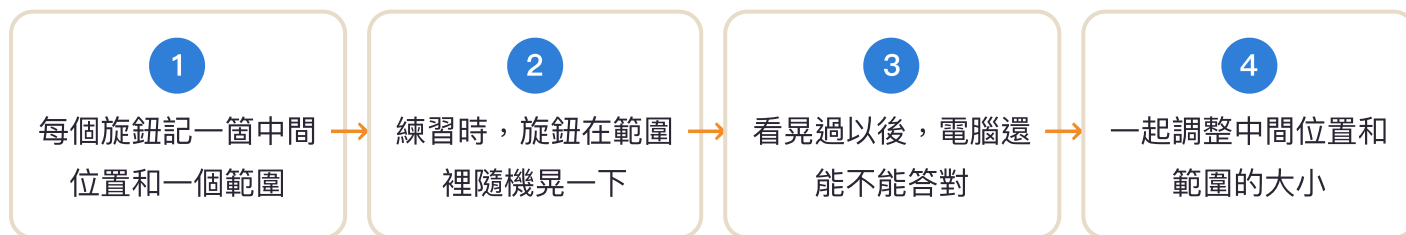
每個旋鈕的位置都要記下來。如果每個都要記成「3.1415926」這樣，描述長度就很長。

記得這麼細，說明電腦把很多小細節都塞進了旋鈕。其中不少是碰巧的細節。這就是死記硬背。

5 Hinton 的辦法：模糊旋鈕

菜譜甲寫「鹽 3.274 克」，菜譜乙寫「鹽一小勺左右」。照著菜譜乙做，菜也很好吃，說明什麼？說明你掌握了這道菜的**真正竅門**，不靠精確的數字。

Hinton 和 van Camp 就是這樣教電腦的。他們讓每個旋鈕不記一個準確的數，只記一個**大概範圍**。練習時，每個旋鈕都在自己的範圍裡隨便晃一晃。科學家把這種隨機的晃動叫**噪聲**。



↻ 換一道練習題，重複很多次

範圍越大，旋鈕越模糊，要記的東西就越少。可是範圍太大，電腦會答錯。所以電腦要找一個平衡：**範圍儘量大，答案還要對**。

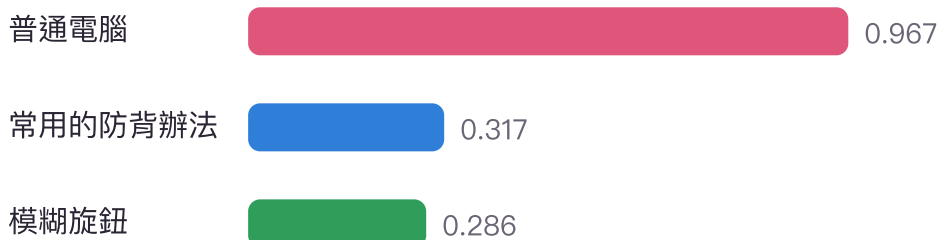
旋鈕們自己分出了重要和不重要

練完以後，有的旋鈕範圍很寬，說明它不太重要。有的旋鈕範圍很窄，說明它必須擰得準。旋鈕晃來晃去，電腦還能答對，說明它學到的是**真規律**，不是死記的細節。

6 論文裡的實驗：105 道練習題

作者找了一個很難的任務：看一種分子的信息，預測它的「活性」有多強。這種分子叫多肽，可以想成小號的蛋白質。難在哪裡？練習題只有 **105 道**，電腦裡的旋鈕卻有 **521 個**。旋鈕比練習題多得多，電腦非常容易死記硬背。

作者另外準備了 420 道新題，用來考試。下面是 3 種電腦的考試成績。橫條越短，錯得越少：



數字 1 表示和「每次都猜平均數」一樣差。普通電腦接近 1，它把練習題背下來了。當時常用的防背辦法也有效，不過要科學家仔細調好。模糊旋鈕錯得最少。

作者也老實說了：這只是初步實驗，還要和更多方法比較。這篇論文最珍貴的，是它的**想法**。

7 這和今天有什麼關係

這節課的主線是一句話：**壓縮就是理解**。「壓縮」就是把一件事說得更短，又不丟掉要緊的內容。能說得短，說明你找到了規律。

- **和第 1 課連起來**：聊天 AI 靠「猜下一個字」練出來。猜得越準，就越能把文章壓得越短。「會猜」和「會壓縮」是同一種本事。
- **旋鈕不用太精確**：今天的大型 AI 有幾十億個旋鈕。工程師常把旋鈕記得粗一點，讓 AI 變小，AI 照樣能用。
- **誠實地說**：今天的大型 AI 一般不用這篇論文的具體做法。它們主要靠讀海量資料來防止死記硬背。

第 5 課和第 26 課還會接著講「壓縮就是理解」。

8 動手：傳話畫圖

玩法（2—4 人，約 15 分鐘）

1. 一人當「描述員」，只有他看下面的圖。其他人當「畫家」，先遮住圖。
2. 描述員看圖 1，用盡量少的字說出來，不能比畫。另一人數他說了幾個字。
3. 畫家照著畫。對照原圖，數 5 個要點對了幾個：形狀、數量、位置、大小、方向。再換圖 2、圖 3。

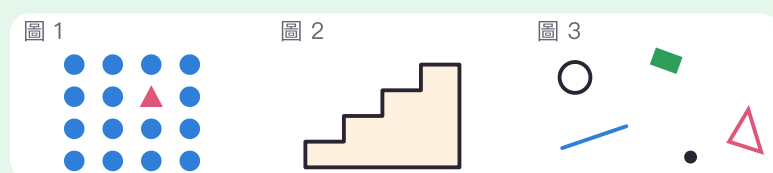


	圖 1	圖 2	圖 3
字數			
對了幾個			

你會發現：圖 1、圖 2 有規律，字少，畫得也準。圖 3 沒有規律，字最多，還最容易畫錯。找到規律，才能用最短的話傳過去。

9 動手：大概就行

玩法（1—3 人，約 10 分鐘）

1. 剪一條 30 釐米長的紙條，只在 0、10、20、30 釐米處畫線。這就是「模糊尺子」。
2. 只用模糊尺子完成下面的任務，判斷它夠不夠用。不夠用的，再拿真尺子量。

任務	夠用嗎（✓ 或 ✗）
這本課本能不能平放進書包？	
這支鉛筆能不能放進鉛筆盒？	
兩支差不多長的鉛筆，哪支更長？	
自己想一個任務： _____	

你會發現：大部分任務，知道個大概就夠了，只有少數要量得很準。Hinton 的電腦也一樣：多數旋鈕可以很模糊，少數要擰得準。

10 詞語小卡片

描述長度

說清楚一件事要用多少字。好的說法 = 短規則 + 少例外，加起來越短越好。

過擬合（死記硬背）

電腦把練習題連同碰巧的小細節都背下來，遇到新題就答錯。

權重（旋鈕）

電腦裡成千上萬個可以調的數。它們一起決定電腦怎麼回答。

模糊旋鈕

只記一個大概範圍的旋鈕。練習時它會隨機晃動，逼電腦去學真規律。

11 想一想

1. 你背過的東西里，哪些是「死記」的，哪些是「找到規律」的？找到規律的那些，是不是更難忘？

2. 旋鈕的範圍太大，電腦會答錯；範圍太小，電腦會死記硬背。為什麼要在中間找平衡？

3. 請用最短的話，把今天這節課講給家人聽。你用了幾個字？

一句話記住：能用最短的話說清楚，才是真的懂了。

給家長和老師：本課講的是 Geoffrey Hinton 和 Drew van Camp 1993 年的論文 [Keeping Neural Networks Simple by Minimizing the Description Length of the Weights](#) (COLT 1993)。論文用最小描述長度 (MDL) 原則約束神經網絡：給每個權重一個可學習的高斯噪聲，權重越「模糊」，攜帶的信息越少，越不容易過擬合。講義裡的「旋鈕」就是權重，「範圍」對應噪聲的大小。講義略去了公式、KL 散度和比特回收 (bits back) 論證。

實驗數據來自論文：多肽活性預測任務，105 個訓練樣本，420 個測試樣本，網絡有 521 個權重；圖中數字是相對誤差（1.0 相當於總猜均值），「常用的防背辦法」指仔細調過係數的權重衰減。講義中的熱身圖形串、水杯報告、菜譜、3.1415926 旋鈕、傳話畫圖的 3 幅圖都是示意，不是原文內容。「把旋鈕記得粗一點」指今天大模型的量化 (quantization)，這是後來的工程做法，不是本文方法。

想深入可看本站的 [第 4 天中文導讀](#)，代碼見開源項目 [30u30 第 4 天目錄](#) (英文)。