

最短的說法贏了

本課講 2004 年的一篇文章。科學家 Peter Grünwald 寫了一份入門講解，介紹「最小描述長度」這個想法：誰能把事情說得最短，誰就找到了真正的規律。

適合 10 歲左右

約 40 分鐘

需要：鉛筆、一枚硬幣、一位同伴

1 熱身：說短一點

你要在電話裡把下面的東西告訴朋友。朋友得一字不差地寫下來。在橫線上寫一個更短的說法。

1111111111 _____

紅黃紅黃紅黃紅黃紅黃

5, 10, 15, 20, ..., 50

8, 3, 6, 1, 9, 2 _____

前 3 個很好辦，比如第 1 個可以說「10 個 1」。最後一個大概只能一個一個念。為什麼有的能說短，有的不能？這就是今天的主題。

2 規律就是能壓縮的東西

用更短的話說同一件事，叫**壓縮**。壓縮有一個硬規矩：對方聽了短說法，要能**一字不差地寫回原樣**。說法有多長，我們數字數：每個漢字、字母、數字算 1 個字，標點和空格不算。

原來的東西	原來幾個字	短說法	短說法幾個字
AAAAAAAAAA	10	10個A	4
2, 4, 6, 8, ..., 98, 100（一共 50 個數）	97	從2開始每次加2到100	12

能壓縮，是因為東西里藏著**規律**。找到了規律，就能用一句話代替一長串內容。反過來說：一樣東西能壓縮，就說明它裡面有規律。

3 擲硬幣：壓縮不了

再看兩串硬幣的正面和反面。第一串是故意排的。第二串是擲 12 次硬幣記下來的（示意）。

第一串	正 反 正 反 正 反 正 反 正 反 正 反	「正反重複6次」，6 個字
第二串	正 正 反 正 反 反 反 正 正 反 正 反	試試看，很難比 12 個字更短

擲硬幣的結果是**隨機**的。隨機就是完全沒有規律，下一次是正是反，誰也猜不到。沒有規律，就沒有東西可以拿來代替。所以隨機的串，最短的說法幾乎就是它自己。偶爾會碰巧出現一小段，比如連著 3 個「反」，可是沒有能一直用下去的規則。

4 規則 + 例外：選最短的

真實的東西常常是「大部分有規律，偶爾有例外」。看這串數，一共 50 個數：

2, 4, 6, ..., 46, 48, **51**, 52, 54, ..., 98, 100

幾乎都是每次加 2，只有第 25 個數是 51，不是 50。我們比一比 3 種說法。每種都分成兩部分：**規則**，加上規則管不到的**例外**。

說法	規則	例外（規則沒說清的部分）	一共
A. 全背下來	寫全 50 個數（97 字）	沒有（0 字）	97 字
B. 太簡單	「都是數字」（4 字）	沒幫上忙，還得寫全 50 個數（97 字）	101 字
C. 剛剛好	「從2開始每次加2到100」（12 字）	「第25個是51」（7 字）	19 字

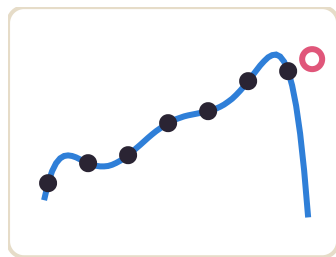


C 最短，所以選 C。這就是**最小描述長度原則**，英文縮寫是 MDL。「描述」就是說法，「長度」就是字數，「最小」就是最短。合起來是：**在幾種解釋裡，選「規則 + 例外」加起來最短的那個**。

注意，例外不能省。省掉它，對方會寫成 50，那就錯了。這個想法最早由科學家 Jorma Rissanen 在 1978 年提出，Grünwald 的講解把它從頭講清楚了。

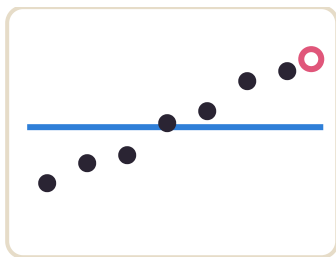
5 不死記，也不偷懶

A 像死記硬背的同學，B 像偷懶的同學。MDL 在兩者中間找平衡。Grünwald 的講解用「畫一條線穿過一堆點」做例子。下面是我們的示意：7 個點大致排成斜線，有點歪。歪是小誤差，比如尺子沒對準。



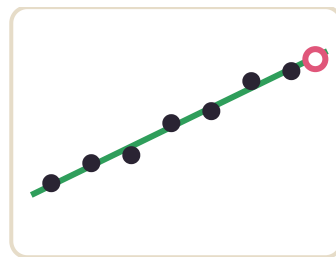
A 全背下來

線彎來彎去，穿過每個點



B 太簡單

一條平線，離很多點都遠



C 剛剛好

一條斜直線，加一點小誤差

粉色空心圈是後來新量到的點。A 的線在最右邊猛地往下拐，B 的平線太低，只有 C 的斜線從它旁邊經過。

A 把偶然的小誤差也當成了規律。科學家把這種毛病叫**過擬合**，意思是「背得太死，換一道新題就不會了」。MDL 會把小誤差放進「例外」，不讓它們混進規則。所以說，**MDL 是防止死記硬背的數學辦法**。

和第 4 課連起來

第 4 課講 Hinton 和 van Camp 1993 年的文章。他們把這個想法用在神經網絡上：網絡本身算「規則」，網絡猜錯的地方算「例外」，兩樣加起來越短越好。這一課講的是這個想法本身。

6 這和今天有什麼關係

第 1 課講過：ChatGPT、Claude 這些聊天 AI，都是靠「猜下一個字」練出來的。會猜和會壓縮，其實連在一起。

會猜的人，傳話更省字（示意）

你給一位很會猜的朋友傳一句話，朋友一個字一個字地搶先猜。猜對了，你什麼都不用說；猜錯了，你才告訴他正確的字。朋友越會猜，你說的字越少。他的猜測就是規則，猜錯的地方就是例外。

聊天 AI 也一樣。它猜下一個字猜得越準，用它壓縮文字就越短。所以很多科學家說：學習和壓縮是同一件事的兩面。

不過還有難題沒解決。聊天 AI 本身非常大，「規則」很長，按 MDL 算應該很吃虧，可它在新問題上表現不錯。怎樣量出大網絡的「規則」有多長，科學家還在研究。

7 動手：壓縮挑戰

玩法（1—2 個人，約 12 分鐘）

1. 擲 12 次硬幣，按順序記下「正」和「反」，填在最後一行。
2. 每行想一個更短的寫法，數數幾個字（括號也算）。只把短寫法給同伴看，同伴能一字不差地寫回原串，才打 ✓。

原串	字數	我的短寫法	字數	還原
AAAAABBBCCCC	12	5A3B4C（示範）	6	✓
AAAAAAAAAAAAAAAAAAAA	20			
ABCABCABCABCABC	15			
BCAACBABCCBA	12			
我擲的硬幣：	12			

你會發現：有規律的串能壓得很短。最後兩行沒有規律，幾乎壓不短。

8 動手：找規律比賽

玩法（2—4 個人，約 15 分鐘）

1. 每人用「規則 + 例外」描述下面的數串，再數一數一共幾個字。
2. 交換紙張，同伴照著寫出整串數。寫錯一個數，這一行不算分。算分的人裡字數最少的得 1 顆星。

數串	規則	例外	字數	★
1, 2, 4, 8, 16, 32, 64, 128				
5, 5, 5, 5, 5, 5, 5, 5, 9, 5, 5, 5				
10, 20, 30, 40, 51, 60, 70, 80, 90, 100				
7, 2, 9, 4, 4, 1, 8, 3				

你會發現：規則太囉嗦，字數多；規則太含糊，例外一大堆。最好的說法在中間。最後一行最短的說法，就是照抄一遍。

9 詞語小卡片

壓縮

用更短的話說同一件事，而且對方能一字不差地寫回原樣。

隨機

完全沒有規律，下一個是什麼誰也猜不到。隨機的東西壓縮不了。

最小描述長度 (MDL)

在幾種解釋裡，選「規則 + 例外」加起來最短的那個。

過擬合 (死記硬背)

把偶然的小誤差也當成規律背下來，遇到新情況就猜錯。

10 想一想

1. 你一週的課程表能壓縮嗎？用「規則 + 例外」寫一個短說法，再數數有幾個字。

2. 有人說：「所有數串都能用 4 個字說完，它們都是數字。」這為什麼不算一個好說法？

3. 一位同學把練習冊的答案全背了下來，另一位同學找到了解題方法。考試出了新題，誰更可能做對？用這一課的話說說為什麼。

一句話記住：規則加例外，總長最短的解釋最好。能壓縮，就說明找到了規律。

給家長和老師：本課講的是 Peter Grünwald 2004 年的教程 [A Tutorial Introduction to the Minimum Description Length Principle](#) (2005 年收入 MIT Press 文集 Advances in Minimum Description Length)。MDL 由 Jorma Rissanen 1978 年提出。教程從兩部分編碼講到精煉 MDL 的通用編碼，並用多項式擬合說明怎樣在擬合度與複雜度之間取捨。講義裡的「規則」對應模型，「例外」對應藉助模型描述數據的部分。

「按字數量長度」是為孩子定的簡化數法，真正的 MDL 用比特計算編碼長度。熱身題、「第 25 個是 51」、硬幣序列、畫線圖、「會猜的朋友」和兩個活動的數串都是示意，不是原文內容。想深入可看本站的 [第 5 天中文導讀](#)，代碼見開源項目 [30u30 第 5 天目錄](#) (英文)。

極客教育 jikejiao.com 原創講義 · 2026-10