

最短的说法赢了

本课讲 2004 年的一篇文章。科学家 Peter Grünwald 写了一份入门讲解，介绍“最小描述长度”这个想法：谁能把事情说得最短，谁就找到了真正的规律。

适合 10 岁左右

约 40 分钟

需要：铅笔、一枚硬币、一位同伴

1 热身：说短一点

你要在电话里把下面的东西告诉朋友。朋友得一字不差地写下来。在横线上写一个更短的说法。

1111111111 _____

红黄红黄红黄红黄红黄

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

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

前 3 个很好办，比如第 1 个可以说“10 个 1”。最后一个大概只能一个一个念。为什么有的能说短，有的不能？这就是今天的主题。

2 规律就是能压缩的东西

用更短的话说同一件事，叫**压缩**。压缩有一个硬规矩：对方听了短说法，要能**一字不差地写回原样**。说法有多长，我们数字数：每个汉字、字母、数字算 1 个字，标点和空格不算。

原来的东西	原来几个字	短说法	短说法几个字
AAAAAAAAAA	10	10个A	4
2, 4, 6, 8, ..., 98, 100（一共 50 个数）	97	从2开始每次加2到100	12

能压缩，是因为东西里藏着**规律**。找到了规律，就能用一句话代替一长串内容。反过来说：一样东西能压缩，就说明它里面有规律。

3 掷硬币：压缩不了

再看两串硬币的正面和反面。第一串是故意排的。第二串是掷 12 次硬币记下来的（示意）。

第一串	正 反 正 反 正 反 正 反 正 反 正 反	“正反重复6次”，6 个字
第二串	正 正 反 正 反 反 反 正 正 反 正 反	试试看，很难比 12 个字更短

掷硬币的结果是**随机**的。随机就是完全没有规律，下一次是正是反，谁也猜不到。没有规律，就没有东西可以拿来代替。所以随机的串，最短的说法几乎就是它自己。偶尔会碰巧出现一小段，比如连着 3 个“反”，可是没有能一直用下去的规则。

4 规则 + 例外：选最短的

真实的东西常常是“大部分有规律，偶尔有例外”。看这串数，一共 50 个数：

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

几乎都是每次加 2，只有第 25 个数是 51，不是 50。我们比一比 3 种说法。每种都分成两部分：**规则**，加上规则管不到的**例外**。

说法	规则	例外（规则没说清的部分）	一共
A. 全背下来	写全 50 个数（97 字）	没有（0 字）	97 字
B. 太简单	“都是数字”（4 字）	没帮上忙，还得写全 50 个数（97 字）	101 字
C. 刚刚好	“从2开始每次加2到100”（12 字）	“第25个是51”（7 字）	19 字

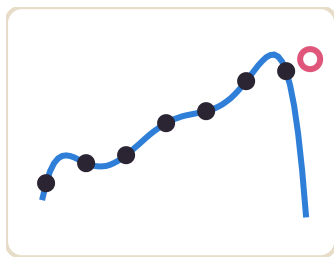


C 最短，所以选 C。这就是**最小描述长度原则**，英文缩写是 MDL。“描述”就是说法，“长度”就是字数，“最小”就是最短。合起来是：**在几种解释里，选“规则 + 例外”加起来最短的那个**。

注意，例外不能省。省掉它，对方会写成 50，那就错了。这个想法最早由科学家 Jorma Rissanen 在 1978 年提出，Grünwald 的讲解把它从头讲清楚了。

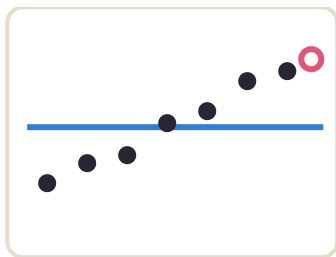
5 不死记，也不偷懒

A 像死记硬背的同学，B 像偷懒的同学。MDL 在两者中间找平衡。Grünwald 的讲解用“画一条线穿过一堆点”做例子。下面是我们的示意：7 个点大致排成斜线，有点歪。歪是小误差，比如尺子没对准。



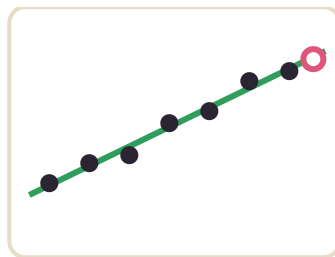
A 全背下来

线弯来弯去，穿过每个点



B 太简单

一条平线，离很多点都远



C 刚刚好

一条斜直线，加一点小误差

粉色空心圈是后来新量到的点。A 的线在最右边猛地往下拐，B 的平线太低，只有 C 的斜线从它旁边经过。

A 把偶然的小误差也当成了规律。科学家把这种毛病叫**过拟合**，意思是“背得太死，换一道新题就不会了”。MDL 会把小误差放进“例外”，不让它们混进规则。所以说，**MDL 是防止死记硬背的数学办法**。

和第 4 课连起来

第 4 课讲 Hinton 和 van Camp 1993 年的文章。他们把这个想法用在神经网络上：网络本身算“规则”，网络猜错的地方算“例外”，两样加起来越短越好。这一课讲的是这个想法本身。

6 这和今天有什么关系

第 1 课讲过：ChatGPT、Claude 这些聊天 AI，都是靠“猜下一个字”练出来的。会猜和会压缩，其实连在一起。

会猜的人，传话更省字（示意）

你给一位很会猜的朋友传一句话，朋友一个字一个字地抢先猜。猜对了，你什么都不用说；猜错了，你才告诉他正确的字。朋友越会猜，你说的字越少。他的猜测就是规则，猜错的地方就是例外。

聊天 AI 也一样。它猜下一个字猜得越准，用它压缩文字就越短。所以很多科学家说：学习和压缩是同一件事的两面。

不过还有难题没解决。聊天 AI 本身非常大，“规则”很长，按 MDL 算应该很吃亏，可它在新问题上表现不错。怎样量出大网络的“规则”有多长，科学家还在研究。

7 动手：压缩挑战

玩法（1—2 个人，约 12 分钟）

- 1. 掷 12 次硬币，按顺序记下“正”和“反”，填在最后一行。
- 2. 每行想一个更短的写法，数数几个字（括号也算）。只把短写法给同伴看，同伴能一字不差地写回原串，才打 ✓。

原串	字数	我的短写法	字数	还原
AAAAABBBCCCC	12	5A3B4C（示范）	6	✓
AAAAAAAAAAAAAAAAAAAA	20			
ABCABCABCABCABC	15			
BCAACBABCCBA	12			
我掷的硬币：	12			

你会发现：有规律的串能压得很短。最后两行没有规律，几乎压不短。

8 动手：找规律比赛

玩法（2—4 个人，约 15 分钟）

- 1. 每人用“规则 + 例外”描述下面的数串，再数一数一共几个字。
- 2. 交换纸张，同伴照着写出整串数。写错一个数，这一行不算分。算分的人里字数最少的得 1 颗星。

数串	规则	例外	字数	★
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				

你会发现：规则太啰嗦，字数多；规则太含糊，例外一大堆。最好的说法在中间。最后一行最短的说法，就是照抄一遍。

9 词语小卡片

压缩

用更短的话说同一件事，而且对方能一字不差地写回原样。

随机

完全没有规律，下一个是什么谁也猜不到。随机的东西压缩不了。

最小描述长度 (MDL)

在几种解释里，选“规则 + 例外”加起来最短的那个。

过拟合 (死记硬背)

把偶然的小误差也当成规律背下来，遇到新情况就猜错。

10 想一想

1. 你一周的课程表能压缩吗？用“规则 + 例外”写一个短说法，再数数有几个字。

2. 有人说：“所有数串都能用 4 个字说完，它们都是数字。”这为什么不算一个好说法？

3. 一位同学把练习册的答案全背了下来，另一位同学找到了解题方法。考试出了新题，谁更可能做对？用这一课的话说说为什么。

一句话记住：规则加例外，总长最短的解释最好。能压缩，就说明找到了规律。

给家长和老师：本课讲的是 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。教程从两部分编码讲到精炼 MDL 的通用编码，并用多项式拟合说明怎样在拟合度与复杂度之间取舍。讲义里的“规则”对应模型，“例外”对应借助模型描述数据的部分。

“按字数量长度”是为孩子定的简化数法，真正的 MDL 用比特计算编码长度。热身题、“第 25 个是 51”、硬币序列、画线图、“会猜的朋友”和两个活动的数串都是示意，不是原文内容。想深入可看本站的 [第 5 天中文导读](#)，代码见开源项目 [30u30 第 5 天目录](#) (英文)。