

第 4 课

说得越短，懂得越多

本课讲 1993 年的一篇论文。作者是 Hinton 和 van Camp。他们让电脑的每个“旋钮”只记个大概，电脑反而学得更好。

适合 10 岁左右

约 40 分钟

需要：铅笔、白纸、一位同伴

1 热身：用最少的字说清楚

下面每一行都是一串图形。请你用最少的字，把它说清楚。别人听了你的话，要能一个不差地写出来。

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

① 很好说：“10 个圆”。② 也行：“圆三角，重复 5 次”。③ 要多说一点：“10 个圆，第 7 个换成三角”。④ 呢？你找不到规律，只能一个一个念。

发现了吗？**有规律的东西，可以说得很短。**没有规律的东西，只能一个一个死记。

2 好的学习 = 用最短的话说清楚

老师想知道：今天班上 30 个同学，谁带了水杯？班长可以用 3 种方法报告。

方法	班长怎么说	说了多长
A. 全背下来	“张一带了，李二带了……”念 30 个名字	很长
B. 规则太粗	“没人带。除了张一，李二……”念 28 个名字	还是很长
C. 好规则	“都带了，只有小明和小红没带。”	很短

方法 C = 一条短规则（都带了）+ 几个例外（小明、小红）。规则越短，例外越少，加起来就越短。加起来最短的说法，往往就是**最懂的说法**。

好规则还能猜新情况。明天来了一个转学生，方法 C 的班长会猜：“他大概也带了。”方法 A 的班长只背过 30 个名字，没法猜。

说清楚一件事要用多少字，科学家叫它**描述长度**。这篇论文的想法是：电脑学东西时，也要让描述长度越短越好。

3 死记硬背的电脑

你见过只背答案的同学吗？他背下“第 1 题选 B，第 2 题选 C”。练习册上的题，他全做对了。可是考试换了新题，他就不会了。

电脑也会犯这个毛病。它把练习题的答案全背下来，连题目里碰巧出现的小细节也背下来。练习时它满分，遇到新题却错一大片。这个毛病叫**过拟合**。你可以把它记成“死记硬背”。

能记的东西多，练习题少，最容易死记硬背。怎么防止？Hinton 的答案是：**别让它记得太精确**。要明白这句话，先认识电脑里的“旋钮”。

4 电脑里的旋钮

第 1 课说过：电脑猜错了，就“悄悄调整自己”。它调的是什么？是成千上万个**旋钮**。

旋钮就像收音机上调音量的按钮。每个旋钮拧到一个位置，就是一个数。所有旋钮一起决定电脑怎么回答问题。科学家管这些旋钮叫**权重**。



精确：3.1415926... 模糊：大概 3 左右

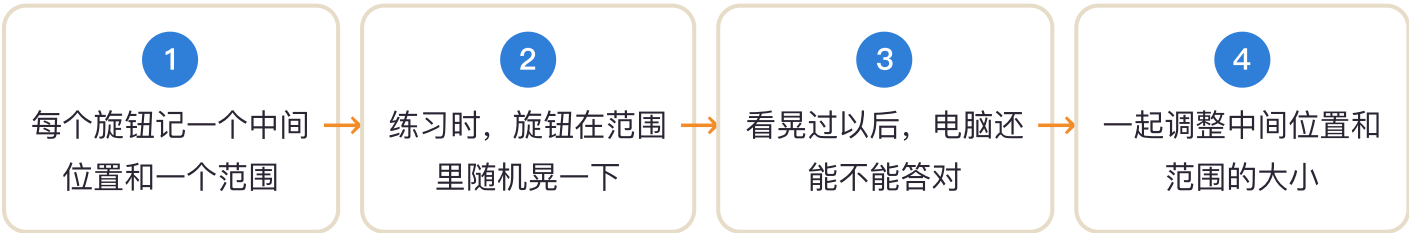
每个旋钮的位置都要记下来。如果每个都要记成“3.1415926”这样，描述长度就很长。

记得这么细，说明电脑把很多小细节都塞进了旋钮。其中不少**是碰巧的细节**。这就是死记硬背。

5 Hinton 的办法：模糊旋钮

菜谱甲写“盐 3.274 克”，菜谱乙写“盐一小勺左右”。照着菜谱乙做，菜也很好吃，说明什么？说明你掌握了这道菜的**真正窍门**，不靠精确的数字。

Hinton 和 van Camp 就是这样教电脑的。他们让每个旋钮不记一个准确的数，只记一个**大概范围**。练习时，每个旋钮都在自己的范围里随便晃一晃。科学家把这种随机的晃动叫**噪声**。



🔄 换一道练习题，重复很多次

范围越大，旋钮越模糊，要记的东西就越少。可是范围太大，电脑会答错。所以电脑要找一个平衡：**范围尽量大，答案还要对**。

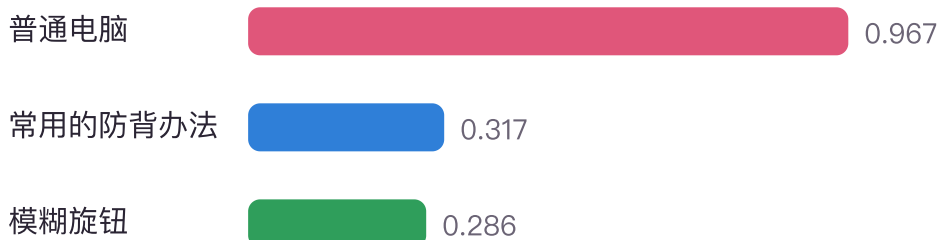
旋钮们自己分出了重要和不重要

练完以后，有的旋钮范围很宽，说明它不太重要。有的旋钮范围很窄，说明它必须拧得准。旋钮晃来晃去，电脑还能答对，说明它学到的是**真规律**，不是死记的细节。

6 论文里的实验：105 道练习题

作者找了一个很难的任务：看一种分子的信息，预测它的“活性”有多强。这种分子叫多肽，可以想成小号的蛋白质。难在哪里？练习题只有 **105 道**，电脑里的旋钮却有 **521 个**。旋钮比练习题多得多，电脑非常容易死记硬背。

作者另外准备了 420 道新题，用来考试。下面是 3 种电脑的考试成绩。横条越短，错得越少：



数字 1 表示和“每次都猜平均数”一样差。普通电脑接近 1，它把练习题背下来了。当时常用的防背办法也有效，不过要科学家仔细调好。模糊旋钮错得最少。

作者也老实说了：这只是初步实验，还要和更多方法比较。这篇论文最珍贵的，是它的**想法**。

7 这和今天有什么关系

这节课的主线是一句话：**压缩就是理解**。“压缩”就是把一件事说得更短，又不丢掉要紧的内容。能说得短，说明你找到了规律。

- **和第 1 课连起来**：聊天 AI 靠“猜下一个字”练出来。猜得越准，就越能把文章压得越短。“会猜”和“会压缩”是同一种本事。
- **旋钮不用太精确**：今天的大型 AI 有几十亿个旋钮。工程师常把旋钮记得粗一点，让 AI 变小，AI 照样能用。
- **诚实地说**：今天的大型 AI 一般不用这篇论文的具体做法。它们主要靠读海量资料来防止死记硬背。

第 5 课和第 26 课还会接着讲“压缩就是理解”。

8 动手：传话画图

玩法（2—4 人，约 15 分钟）

1. 一人当“描述员”，只有他看下面的图。其他人当“画家”，先遮住图。
2. 描述员看图 1，用尽量少的字说出来，不能比画。另一人数他说了几个字。
3. 画家照着画。对照原图，数 5 个要点对了几个：形状、数量、位置、大小、方向。再换图 2、图 3。

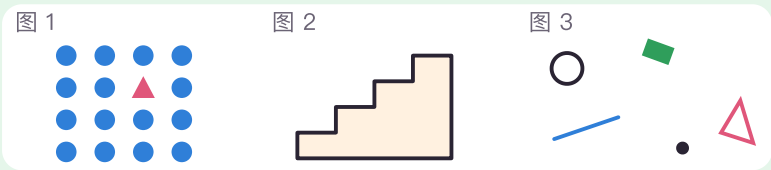


	图 1	图 2	图 3
字数			
对了几个			

你会发现：图 1、图 2 有规律，字少，画得也准。图 3 没有规律，字最多，还最容易画错。找到规律，才能用最短的话传过去。

9 动手：大概就行

玩法（1—3 人，约 10 分钟）

1. 剪一条 30 厘米长的纸条，只在 0、10、20、30 厘米处画线。这就是“模糊尺子”。
2. 只用模糊尺子完成下面的任务，判断它够不够用。不够用的，再拿真尺子量。

任务	够用吗（✓ 或 ✕）
这本课本能不能平放进书包？	
这支铅笔能不能放进铅笔盒？	
两支差不多长的铅笔，哪支更长？	
自己想一个任务： _____	

你会发现：大部分任务，知道个大概就够了，只有少数要量得很准。Hinton 的电脑也一样：多数旋钮可以很模糊，少数要拧得准。

10 词语小卡片

描述长度

说清楚一件事要用多少字。好的说法 = 短规则 + 少例外，加起来越短越好。

过拟合（死记硬背）

电脑把练习题连同碰巧的小细节都背下来，遇到新题就答错。

权重（旋钮）

电脑里成千上万个可以调的数。它们一起决定电脑怎么回答。

模糊旋钮

只记一个大概范围的旋钮。练习时它会随机晃动，逼电脑去学真规律。

11 想一想

1. 你背过的东西里，哪些是“死记”的，哪些是“找到规律”的？找到规律的那些，是不是更难忘？

2. 旋钮的范围太大，电脑会答错；范围太小，电脑会死记硬背。为什么要在中间找平衡？

3. 请用最短的话，把今天这节课讲给家人听。你用了几个字？

一句话记住：能用最短的话说清楚，才是真的懂了。

给家长和老师：本课讲的是 Geoffrey Hinton 和 Drew van Camp 1993 年的论文 [Keeping Neural Networks Simple by Minimizing the Description Length of the Weights](#) (COLT 1993)。论文用最小描述长度 (MDL) 原则约束神经网络：给每个权重一个可学习的高斯噪声，权重越“模糊”，携带的信息越少，越不容易过拟合。讲义里的“旋钮”就是权重，“范围”对应噪声的大小。讲义略去了公式、KL 散度和比特回收 (bits back) 论证。

实验数据来自论文：多肽活性预测任务，105 个训练样本，420 个测试样本，网络有 521 个权重；图中数字是相对误差（1.0 相当于总猜均值），“常用的防背办法”指仔细调过系数的权重衰减。讲义中的热身图形串、水杯报告、菜谱、3.1415926 旋钮、传话画图的 3 幅图都是示意，不是原文内容。“把旋钮记得粗一点”指今天大模型的量化 (quantization)，这是后来的工程做法，不是本文方法。

想深入可看本站的 [第 4 天中文导读](#)，代码见开源项目 [30u30 第 4 天目录](#) (英文)。