磐石计划 01 · A 卷

命题报告（学习路径设计）

版本 v3.0；每卷 21 题、150 分；对象为本次对话中已经接触基础数学语言与深度学习入门内容的学习者。现无实际答卷与追踪数据，本文不报告实测难度、区分度、平均分或“保证掌握率”。

一、任务从筛选改为学习

保留正式试卷外形和四类题型，但题目不是互不相关的检查点。A 卷侧重恢复定义与生成方法，B 卷将同一方法放到新的约束、对象或证明任务中。首次作答的错误用于选择支持，不用于给学习者贴能力标签。

版式参照用户提供的 2025 新课标 I 卷数学 PDF 的居中标题、连续题号、题型说明、普通选项和页脚；解析参照用户提供的 2024 模拟答案册的【答案】【分析】【详解】【点睛】。保留 21 题，不照抄参考卷 19 题结构、原数学题或机构署名。参考文件为版式与解析组织来源，不是本套深度学习理论的依据。

题型	题数	分值	学习功能
单项选择	8	40	对象、定义、短计算
多项选择	3	18	理由、反例与概念分离
填空	4	20	从条件重建表达式
解答	6	72	分步发现、证明与迁移

二、核心知识回收网络

知识链	卷间回收	完成后的可见证据
符号与函数族	A1→B1→A19	能区分集合、元素、函数、函数值及不同参数的同一函数。
矩阵、向量与展平	A3/A4→B3/B4→B14→B20	能给每个对象标形状，并区分计算结果与存储表示。
复合、激活与表达	A2/A8/A12→B2/B16→A20→B17	能从分段函数和几何结构说明非线性的作用。
监督学习与风险	A5/A13→A17→B13→B19	能从记录构造损失，再将经验与总体、固定与选后保证分开。
导数、反传与优化	B5→A18→B8→B18→B21	能写局部导数、完整链式因子与条件化下降证明。
范数、积分与误差	A7/A11/A14→B7/B11→B20	能用不同范数评价误差，证明不等式并构造尖峰反例。
可辨识性与压缩	B1→A19→B12/B14→B20	分别讨论函数唯一、参数唯一、存储数量、近似误差。
稳定性、归纳偏置与算力	A10/A15→B10→A21→B21	知道等变不等于不变、残差不等于收缩、MAC 不等于延迟。

箭头表示建议回访，不表示后题必须依赖前题的答案；正式题面提供必要定义，避免前题做错就无法进入后题。学习顺序可按 A 卷小段订正后进入 B 卷对应题，而非先硬做完两卷。

三、逐题命题说明

1. 模型是“输入＋参数”到输出的映射

已学基础：讲义第 1 章。要生成的证据：能在更换参数个数后重新写映射，并区分空间、元素、函数、函数值。

设计动作：先不代数值，先数“可独立指定的对象”：输入有一个数，参数有两个数；输出仍是一个数。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把参数空间当输出空间；把分号当运算符；把一个预测值与完整函数混为一谈。再使用位置：B1、A13、A19。

2. 函数复合的方向

已学基础：讲义第 1、3 章。要生成的证据：能写出内层函数输出，并把它作为一个整体代入外层。

设计动作：把复合拆成两次输入输出交接，而不是从左边开始算。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把 $\circ$ 当乘法；把中间值 -3 当最终输出；交换复合次序。再使用位置：B2、A18、A20。

3. 转置、内积与输出类型

已学基础：讲义第 2 章。要生成的证据：能不计算就判断输出是标量、向量还是矩阵，再正确运算。

设计动作：先看形状：一行乘一列才能输出一个数；再逐项相乘并求和。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：计算出了三个乘积就停下；认为写了转置只是外观变化，没有改变乘法形状。再使用位置：B3、B4、A19。

4. 矩阵乘向量与逐元素激活

已学基础：讲义第 2、3 章。要生成的证据：能逐步写中间值，且能说明不是矩阵“变成”向量。

设计动作：逐层保留中间状态：线性乘法、加偏置、激活是三件事，不能合并着猜。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把矩阵当作结果；先激活输入再做乘法；偏置维度没有跟随输出。再使用位置：A12、A18、B14。

5. 多重集、指示函数与经验平均

已学基础：讲义第 4、8 章。要生成的证据：能把逻辑判断、求和平均和经验期望三种写法互相转换。

设计动作：先列四条记录，再决定每条记录是否出错；重复记录仍占一次权重。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把多重集去重；把分类器的输出当损失；忘记真实标签参与比较。再使用位置：A17、B13、B19。

6. 最小值与最小化点集合

已学基础：讲义第 1、6 章。要生成的证据：能区分最小值、最优参数、驻点，并保留参数约束。

设计动作：用“平方非负＋什么时候取零”，比直接求导更短，也更容易得到全部最优点。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把 min 的函数值写进 argmin；只找到一个最小点；把导数为零当充分条件。再使用位置：B15、B18。

7. 同一向量的三种大小

已学基础：讲义第 7 章。要生成的证据：能计算三种范数，并用缩放性质解释变式。

设计动作：同一组坐标，分别使用绝对值求和和最大绝对值；两种量不能混用。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把二范数 5 当所有范数；把下标当坐标；对负分量不取绝对值。再使用位置：B4、B11、B20。

8. 从定义推出激活函数恒等式

已学基础：讲义第 3、5 章。要生成的证据：能从定义恢复对称关系，而不是背某个输入处的数值。

设计动作：找成对出现的正负输入，而不是把四个分式通分到一起。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：只在 t=0 验证就宣称恒等；把 Sigmoid 误判为奇函数。再使用位置：A18、B2。

9. 逼近、训练与泛化的三道门

已学基础：讲义第 4、8、10 章。要生成的证据：能给出训练拟合不确定外推的具体反例。

设计动作：区分“存在一个好函数”“算法找到它”“新数据也好”三个逻辑命题。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：认为表达能力越强就自动训练越好；把未提供人工标签误解为不能学习任何东西。再使用位置：A17、B16、B19。

10. 结构假设不是无条件保证

已学基础：讲义第 12 章。要生成的证据：能用等式而不是口号区分等变、不变与因果访问限制。

设计动作：把每个架构名称翻译成一个具体的信息访问或参数共享约束。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：等变与不变混用；认为任何注意力都有同一掩码；把序列信息方向当因果发现。再使用位置：B10、B21。

11. 有界不等于全域可积

已学基础：讲义第 7 章。要生成的证据：能解释范数有限所依赖的区域和指数，而不只计算结果。

设计动作：先看函数非零的区域，再算高度的 q 次方乘以区域长度。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：忘记积分区域；把积分值当范数；把有界与可积当同义词。再使用位置：A14、B7、B17。

12. 仿射层折叠与偏置传播

已学基础：讲义第 2 章。要生成的证据：能推导折叠公式，并用分段斜率说明非线性的作用。

设计动作：把第一层完整代入第二层，权重乘在输入上，也要乘在第一层偏置上。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：只乘权重不乘偏置；把两个偏置直接相加；认为可折叠意味着训练轨迹相同。再使用位置：B14、B16、A18。

13. 从记录构造损失，而非背总公式

已学基础：讲义第 4、6 章。要生成的证据：能从一条记录开始搭建平均损失，再用梯度更新。

设计动作：先独立写两个预测及其误差，最后才平均；平方和非负可直接定位最优点。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：不减真实标签；把预测加起来再平方；漏除样本数；把 argmin 写成最小损失 0。再使用位置：B5、B8、B18。

14. L2 误差与上确界不是同一对象

已学基础：讲义第 7 章。要生成的证据：能区分 sup 与 max，并说明单点改动对积分和逐点值的不同影响。

设计动作：平方积分测整体大小，上确界测最高幅值；后者不要求在定义域内取到。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把“没取到 1”写成上确界不存在；积分忘平方或忘开方；忽略端点定义。再使用位置：B7、B17。

15. 局部误差如何变为几何级数

已学基础：讲义第 11 章。要生成的证据：能从递推写出求和，并把可达上界与实际值区分。

设计动作：先连续展开四步，不提前猜是四倍；每条早期误差还会被后续系数缩放。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：直接乘层数；把不等式写成每次实际必然相等；套级数时漏掉初始项 1。再使用位置：A21、B21。

16. 从一次更新走到感知机间隔界

已学基础：讲义第 3、7 章。要生成的证据：能从更新定义独立推出两界与次数界，并解释每个假设。

设计动作：单个坐标可能反复改变，难以直接控制。题目给了正确分离方向 u，于是比较沿 u 的进展与总长度：一个至少按 k 增长，另一个至多按根号 k 增长。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把检查次数当更新次数；没有在展开范数时使用触发条件；直接除以零次更新数；把训练收敛当总体保证。再使用位置：B16、B20。

17. 经验风险与真实风险分别用谁的权重

已学基础：讲义第 4、8 章。要生成的证据：能将同一错误表转换为两种风险，解释未覆盖事件。

设计动作：模型预测不变，只换权重：同一错误表按观测频数加权是训练风险，按总体概率加权是真实风险。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：混用两套概率；忘记训练中完全没观察到 x=2；最小化点不唯一却只写一个模型。再使用位置：B13、B19。

18. 链式法则怎样组织成反向传播

已学基础：讲义第 5 章。要生成的证据：能从局部导数推出矩阵梯度，且数值更新在同一个旧点计算。

设计动作：不要展开成一个巨大的式子。先求损失对输出的导数，再沿每条边传回去；不同入边共用同一个隐藏误差信号。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：漏 Sigmoid 导数；漏输入 xi；将矩阵梯度写成向量；先改 v 再算 W 梯度导致不再是同一点的同时更新。再使用位置：B5、B18、B21。

19. 函数自由度与参数冗余

已学基础：讲义第 2、9 章。要生成的证据：能分别判断参数可辨识性与预测函数可辨识性。

设计动作：先展开二次型，观察 u 和 v 以什么组合出现。样本能约束函数系数，却永远看不到对输出无贡献的方向。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把 4 个存储数当 4 个可辨识函数系数；认为函数相同必然参数相同；漏掉矩阵反对称部分。再使用位置：B1、B12、B20。

20. 从复合到指数增长的折点

已学基础：讲义第 3、10 章。要生成的证据：能自行维护归纳不变量，且知道本题只是精确表示的教学特例。

设计动作：归纳时不仅猜段数，还要保留可继续使用的性质：每段单调覆盖 [0,1]。这保证下一轮恰好切成两段。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：只统计新增折点却没排除旧折点消失；把必要下界当最小可达宽度；把个别构造推广为所有任务。再使用位置：B17。

21. 误差传播与计算成本的联合推理

已学基础：讲义第 7、11 章。要生成的证据：能从模型假设推出条件化结论，不把结构名称当稳定性证明。

设计动作：将实际状态与理想状态相减，先得到一个可以反复使用的单步不等式，再用级数汇总；成本则从矩阵形状另外计数。支持先后：H1 指出观察对象；H2 给中间关系；解析给完整证明；最后撤去提示完成同题改动。

防止的假掌握：把上界当实际轨迹；K=1 仍除以 K-1；把 MAC 次数、FLOPs、延迟和任务准确率互当同义词。再使用位置：B20、B21。

四、经典论文不只作背景装饰

经典证明均以新定义题的方式出现：题面补足超出已学内容的前提，学生完成可由现有工具推出的关键部分。不能将原定理答案直接当作可引用结论；若题目要求证明次数界、收敛率或最佳性，须写出中间不等式及依据。

论文来源与改编边界

以下只列本册使用的来源；原文编号与教学题编号不同。正文重新给出全部考查假设。基础恒等式不冒充论文专属定理，教学特例不替代原论文一般结论。

【P1】 Freund, Y.; Schapire, R. E. (1999). Large Margin Classification Using the Perceptron Algorithm. Machine Learning 37, 277–296.

原文定位 / 教学范围：第 3.1 节、Theorem 1；复述 Block / Novikoff 的经典间隔界。

DOI / 标识：10.1023/A:1007662407062

【P2】 Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. (1986). Learning representations by back-propagating errors. Nature 323, 533–536.

原文定位 / 教学范围：公式 (1)–(6)：前向计算、平方损失及链式求导。

DOI / 标识：10.1038/323533a0

【P3】 Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems 2, 303–314.

原文定位 / 教学范围：通用逼近的存在性背景；本资料不声称证明其完整定理。

DOI / 标识：10.1007/BF02551274

【P4】 Telgarsky, M. (2016). Benefits of depth in neural networks. COLT, PMLR 49, 1517–1539.

原文定位 / 教学范围：第 3.3 节、Lemmas 3.8–3.11：三角波及迭代构造；试卷只证精确表示的简化情形。

DOI / 标识：arXiv:1602.04485

【P11】 Vaswani, A., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems 30.

原文定位 / 教学范围：编码器与解码器注意力、位置与因果掩码背景，不在本卷考完整注意力推导。

DOI / 标识：arXiv:1706.03762

五、学习设计依据与使用边界

闭卷重建与延迟再写的设计参考 Karpicke 与 Roediger（2008）关于提取练习的实验；同题更换条件后的迁移参考 Butler（2010）的实验问题。前者使用外语词汇，后者研究事实与概念保持、迁移，均不直接证明本套数学卷对某个人的学习效果。本卷的具体题目安排和下述 1 日 / 3 日 / 7 日回访属于教学设计，不是论文验证的最优时间表。

Karpicke, J. D.; Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. Science 319, 966–968. DOI: 10.1126/science.1152408。

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. Journal of Experimental Psychology: Learning, Memory, and Cognition 36(5), 1118–1133. DOI: 10.1037/a0019902。