满分之后：模型真的学会了吗？

建议 35-45 分钟；先用纸笔预测，再由老师投影实验核对。本地模型课需要电脑运行下载的代码。进阶运算可分两节课。

学习目标

分别统计猫的漏报和非猫的误报，并在调参前封存测试集。

先修与准备

先做感知机和训练数据课；会算分数和百分比。

材料

新学习单、10 张标签卡（1 猫、9 非猫）、三个信封：训练、验证、测试。

课堂顺序（5 / 10 / 10 / 10 分钟，可另留 5-10 分钟讨论）

1. 先猜：模型永远说“非猫”，它在 1 猫 9 非猫的数据上几分？先算总分，再只看猫。

2. 动手：使用第二组 6 个样本：真值 猫猫非非非非，预测 猫非猫非非非。把每一格分到 TP、FN、FP、TN。

3. 观察：先分三个信封。训练用来学权重，验证用来挑门槛或模型；最终测试在选择结束后才打开一次。

4. 挑战：把同一幅原图的复制件分别放训练和测试，讨论为什么不是独立新题；按来源分组后重分。

满分之后：模型真的学会了吗？

做好的例子

永远说非猫：准确率 9/10=90%，但猫的召回率 0/1=0%。第二组 TP=1、FN=1、FP=1、TN=3，准确率 4/6。

理解检查：先让学生独立回答

看过测试结果，再反复调门槛直到测试高分，还能把它当作一次独立期末考试吗？

参考答案与学习单核对

不能；测试信息已参与选择，需要新的独立测试。第二组 precision=1/(1+1)=50%，recall=1/(1+1)=50%；总分 4/6≈66.7%。样本很少，只作练习，不说明真实部署水平。 挑战：9 猫 1 非猫、全报猫，TP=9、FP=1、FN=0、TN=0；非猫召回率为 0/1，或说非猫全部被误报。重复图应按来源整体分在同一集合，不能跨训练和测试。门槛在验证集上选择。

不要把结论讲大了

准确率、精确率和召回率回答不同问题，没有单一指标适合所有场景；分母为 0 时必须标明未定义。

教学调整

需要帮助时：先只做例子，由老师朗读题目，用实物或纸格代替公式。已经掌握时：一次只改一个条件，先记预测，再验证。遇到不一样的答案，先查规则、编码或数据，不先判学生错。

带回家的活动

给家长看 90% 总分，再揭示猫的数量，问还需要哪些统计。

接着学

多层神经网络 / CNN：带着同一把尺子找横线 / 自己的小语言模型：从计数到生成

教师参考（完整链接也在课程页）

scikit-learn: Common pitfalls: https://scikit-learn.org/stable/common_pitfalls.html