



南京大学
NANJING UNIVERSITY

南大 NOW JOIN US 一个好地方 一个读书的

2021安庆二中宣讲

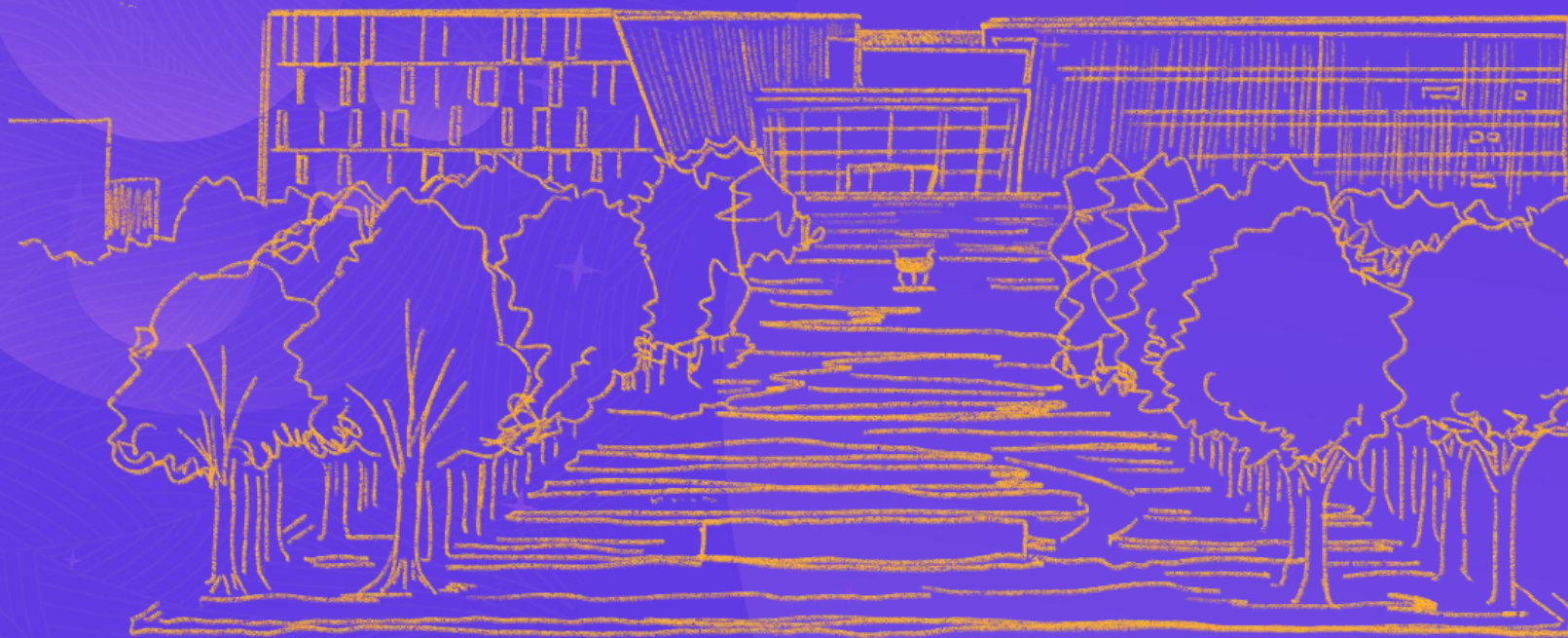
人工智能学院：邢存远



南京大學
NANJING UNIVERSITY

流程：

学校介绍
学科介绍
经验分享





以癌症诊断为例：

病人	性别	年龄	肿瘤尺寸(cm)	肿瘤存在时间(月)	诊断结果
Adam	男	45	21	30	恶性
Bell	男	32	18	12	良性
Carly	女	61	3	7	良性
Dickens	男	53	10	24	恶性
Eason	女	48	7	20	恶性



从机器学习到人的学习

以癌症诊断为例：

病人	性别	年龄	肿瘤尺寸(cm)	肿瘤存在时间(月)	诊断结果
Adam	男	45	21	30	恶性
Bell	男	32	18	12	良性
Carly	女	61	3	7	良性
Dickens	男	53	10	24	恶性
Eason	女	48	7	20	恶性

这时来了一个病人Frank，男，47岁，经检测，肿瘤尺寸5cm，肿瘤已经存在了10个月，如何诊断？



从机器学习到人的学习

目标：

找到一个这样的函数

$result = f(age, gender, size, time)$ → 联想到学过的函数：一元到多元

使预测结果尽可能的精确，即提高准确度



从机器学习到人的学习

目标：

找到一个这样的函数

$result = f(age, gender, size, time)$ → 联想到学过的函数：一元到多元

使预测结果尽可能的精确，即提高准确度

答案 = $f(\text{文本}, \text{数据}, \text{图}) = g(\text{文本}, \text{图})(\text{数据}) = h(\text{数据})$



从机器学习到人的学习

情况一：你在数据库中发现一个叫Friedrich的男人，状况和Frank一样，唯一的差别是比Frank小一岁；

情况二：机器在输入数据后进行计算，不幸的是，在输出时，误将恶性肿瘤概率 p 输出成 $1-p$ ；

情况三：数据库里只有十条病人记录，且很难找到和Frank状况相近的；

情况四：数据库中有不少条数据是错误的（若再加上第一种情况？）

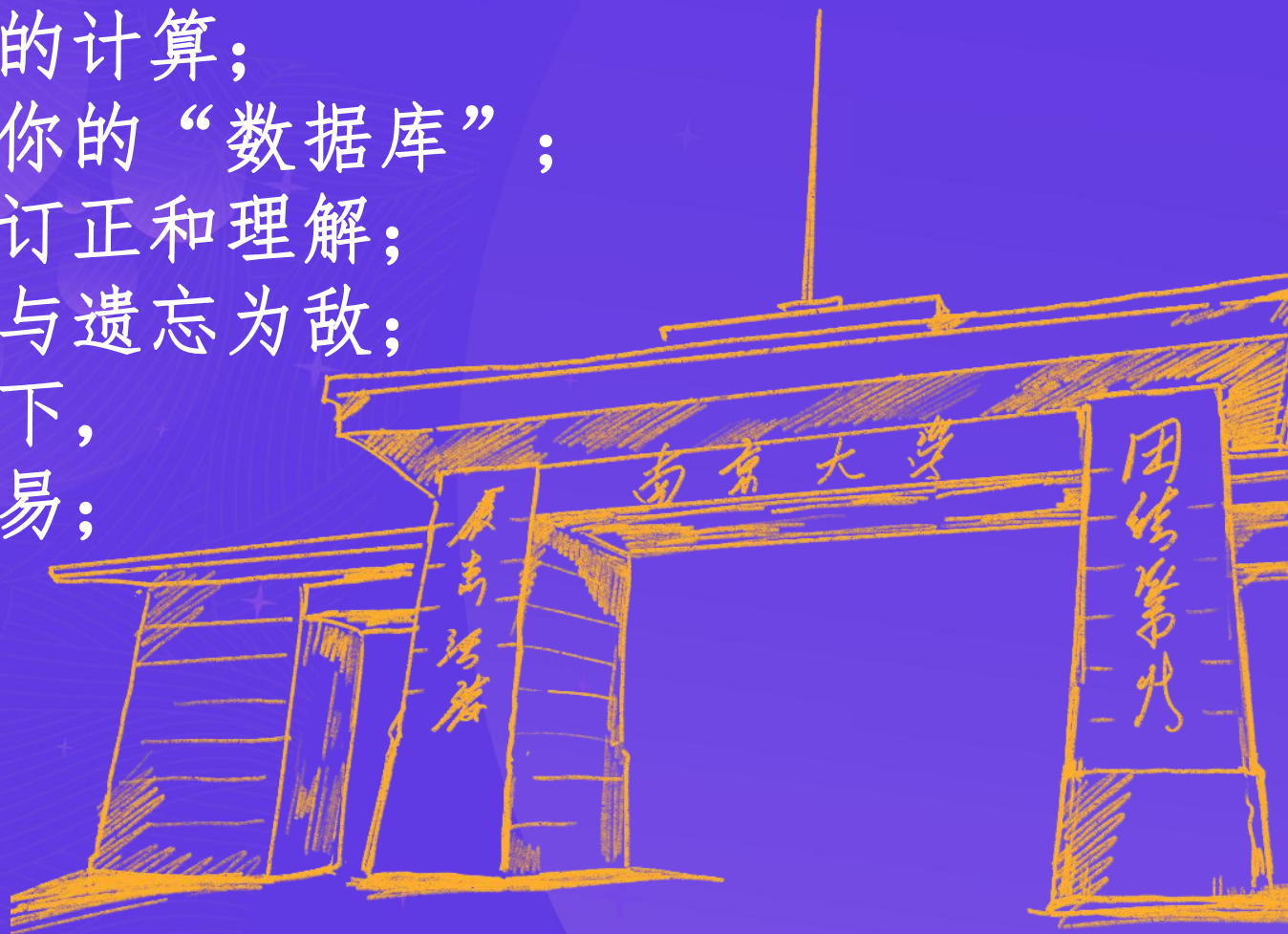
情况五：数据库本应该有10000条数据，但送给学习器后才发现只输入了一半的数据

情况六：你浏览了一遍数据库，发现大部分病人都是20-30岁，导致不少样本都是患良性肿瘤



总结：

- 一，不要忽视最后的计算；
- 二，多做题，充实你的“数据库”；
- 三，做题后要批改订正和理解；
- 四，重视复习，要与遗忘为敌；
- 五，习题难度应上下，
不应太难或太易；
- 六，原题？





从机器学习到人的学习

一些有趣的做法：

- 一，你的同事会将每个月遇到的诊断失误（即学习器的判断错误）的案例会独立保存，不管数据库如何更新，他都不会忘记将这些数据补充输入；
- 二，你的同事A和同事B将各自的学习器整合在一起，诊断时同时输入数据，对这两个输出进行综合考虑。



从机器学习到人的学习

一些有趣的做法：

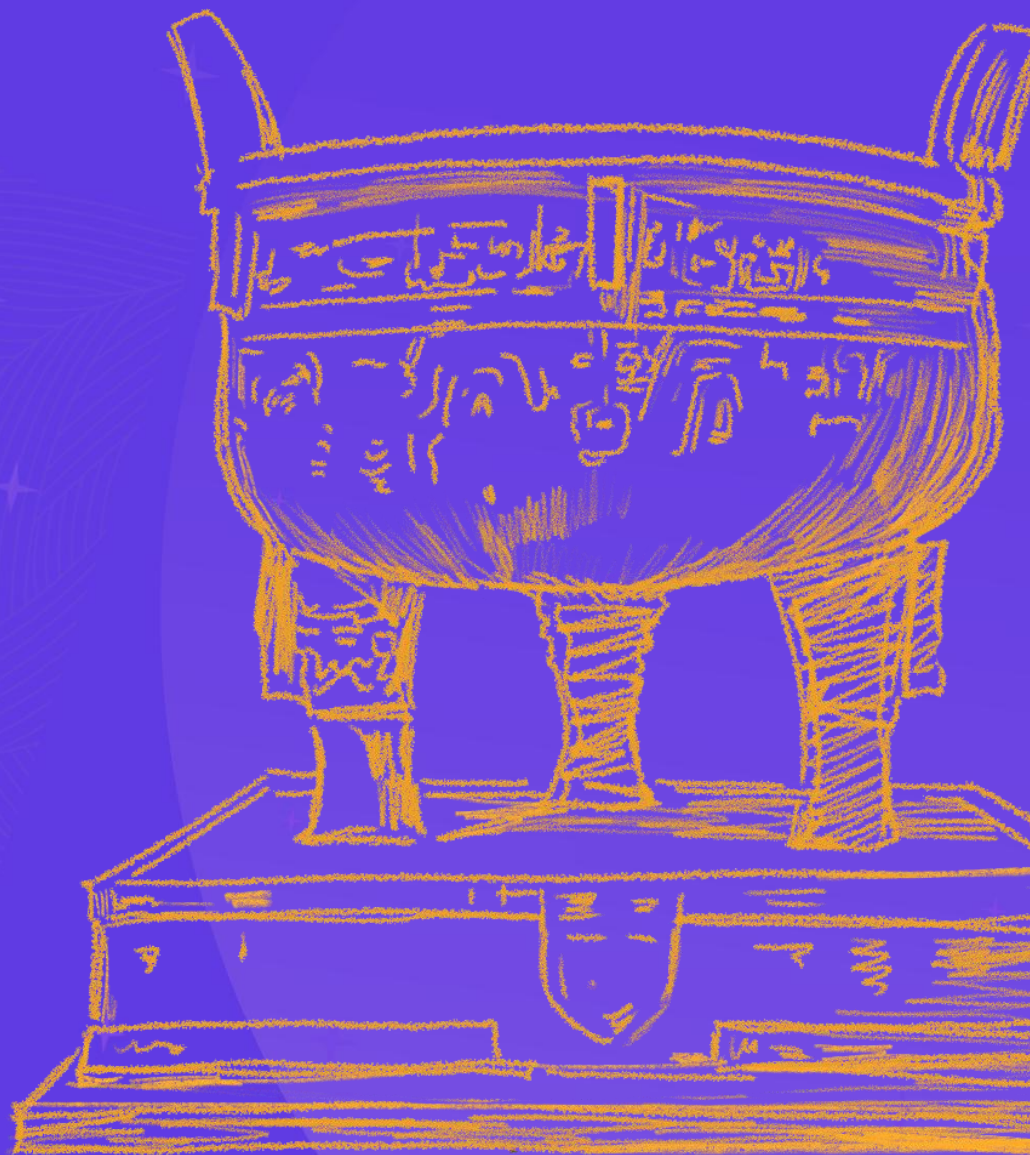
- 一，你的同事会将每个月遇到的诊断失误（即学习器的判断错误）的案例会独立保存，不管数据库如何更新，他都不会忘记将这些数据补充输入；
- 二，你的同事A和同事B将各自的学习器整合在一起，诊断时同时输入数据，对这两个输出进行综合考虑。

错题集+同学间讨论



南京大學
NANJING UNIVERSITY

感谢聆听





南京大學
NANJING UNIVERSITY

南大 NOW JOIN US
一个读书的好地方

南京大学本科招生热线



400-1859680

— 拨我就来帮您 —

