

机器学习：Day2实训内容

李猛 2017-08-10
V1.0.0.0

The World's Local Training Provider

目录

MENU

1

回归

2

非监督学习

3

数据处理

4

模型处理

5

总结Q&A

3.1 数据处理： 处理流程

- 数据处理流程示意图
- Q&A: 现有案例业务讨论



- 数据收集
- 示意图：用户画像

- Q&A: 现有案例业务讨论



3.3 数据处理：数据存储

- 数据存储
- 多种数据来源
- 多种数据格式

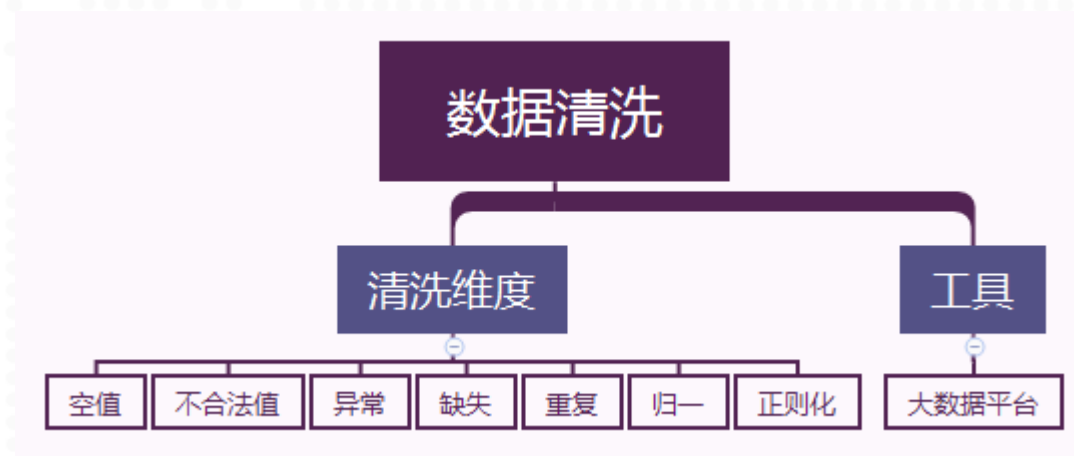
- 示意图：数据平台



- Q&A: 现有案例业务讨论

3.3 数据处理：数据清洗

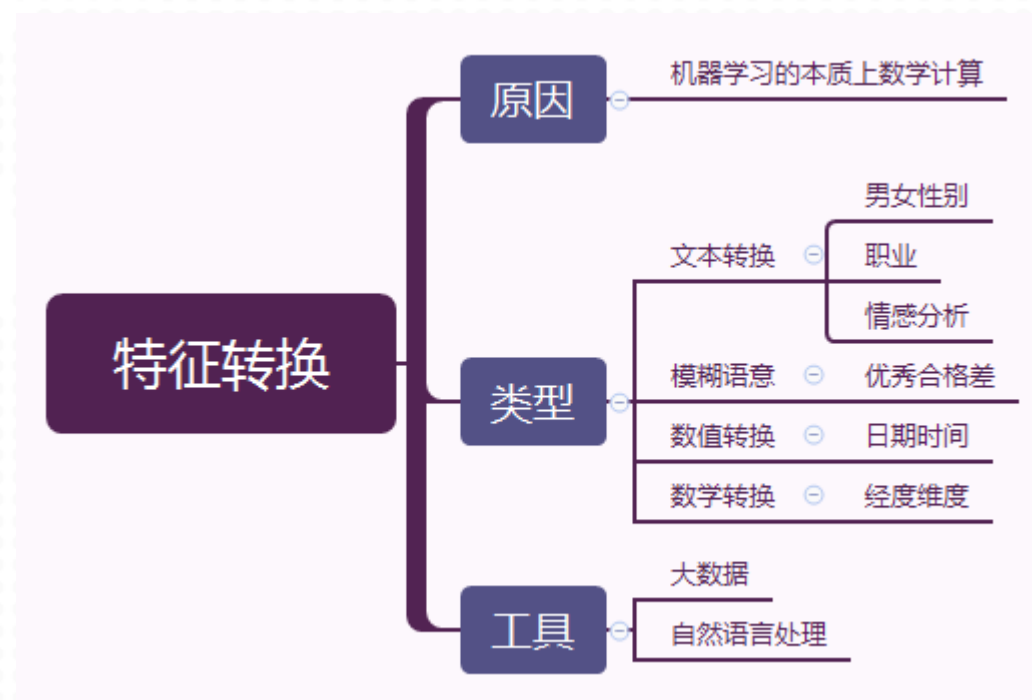
- 数据清洗概念介绍
- 清洗目标：满足机器学习要求
- 示意图：



- Q&A：现有案例业务讨论

3.4 数据处理：特征转换

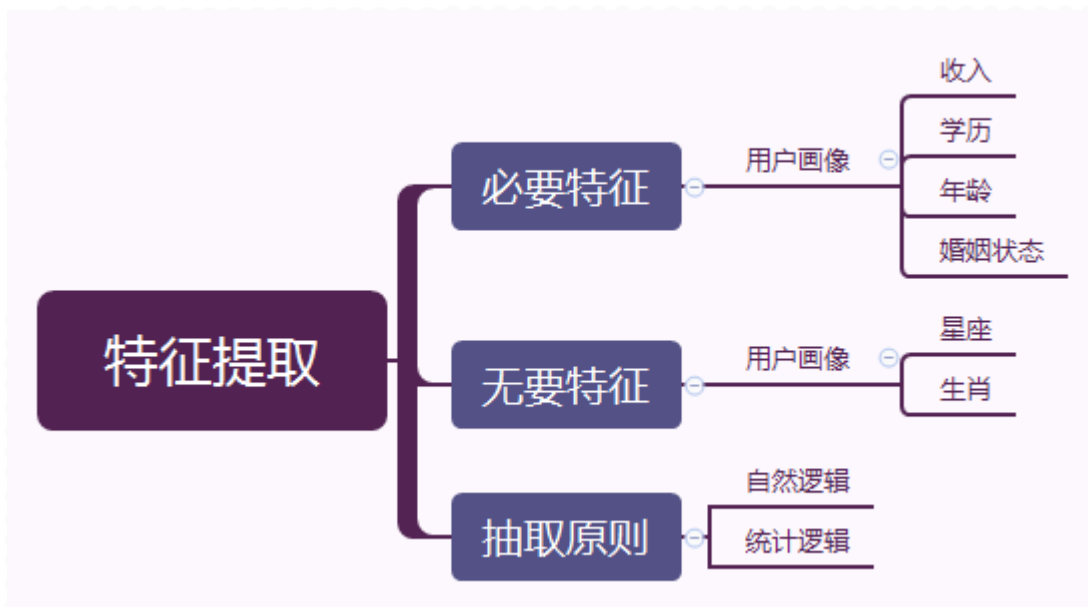
- 特征转换概念介绍
- 转换原因
- 转换类型范围
- 转换工具方法
- 示意图：



- Q&A：现有案例业务讨论

3.5 数据处理：特征提取

- 特征提取概念介绍
- 必要特征
- 无效特征
- 抽取原则
- 示意图：



- Q&A：现有案例业务讨论

3.6 数据处理： 案例练习

- 电影数据MovieLens
- 电影： 去除特征
- 用户： 特征转换
- 电影评分： 去除特征
- 合并电影评分特征（评分+电影特征+观众特征）
- 案例： jupyter
- Q&A： 现有案例业务讨论

4.1 模型处理：模型选择

- 模型选择概念介绍
- 模型目标
- 模型选择
- 模型策略
- 模型算法
- 示意图：



- Q&A：现有案例业务讨论

4.2 模型处理：模型训练

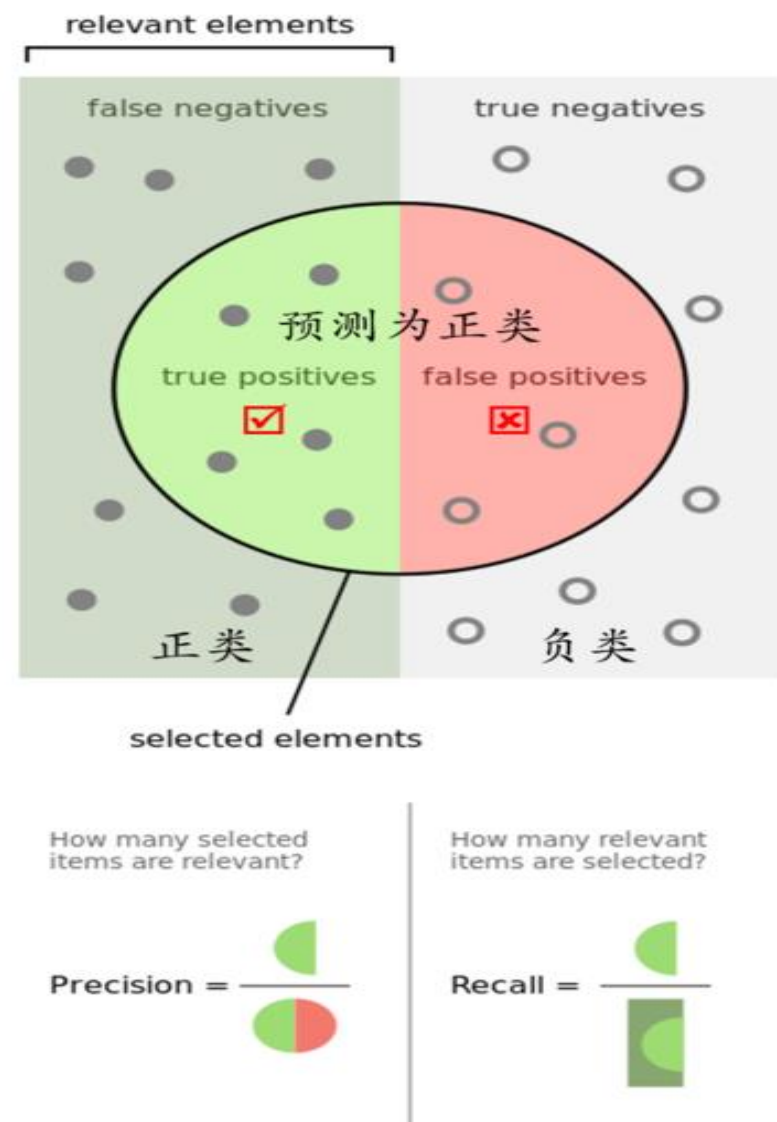
- 模型训练介绍
- 模型交叉训练
- 多种模型并行
- 示意图：
- Q&A：现有案例业务讨论



Baidu 百度

4.3 模型处理：模型评估

- 模型评估介绍
 - 评估方式：交叉评估
 - 评估标准：准确率/召回率
-
- Q&A：现有案例业务讨论



4.3.1 模型处理：模型评估

- 评估标准：ROC/AUC
- 示意图：
- Q&A：现有案例业务讨论

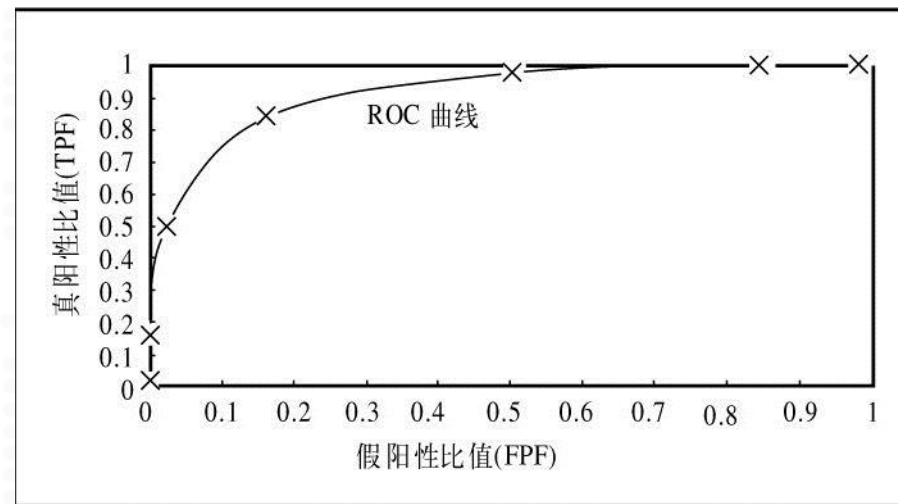


图2 从图1的分布得到的ROC曲线

4.4模型处理：模型迭代

- 模型换选：单一模型到复合模型；线性模型到非线性模型
 - 模型参数：模型训练迭代次数调整
 - 数据特征：更换特征
 - 数据正则化：模型过拟合/模型欠拟合/
 - 数据标准化：数据归一
-
- Q&A：现有案例业务讨论

4.5 模型处理：模型部署

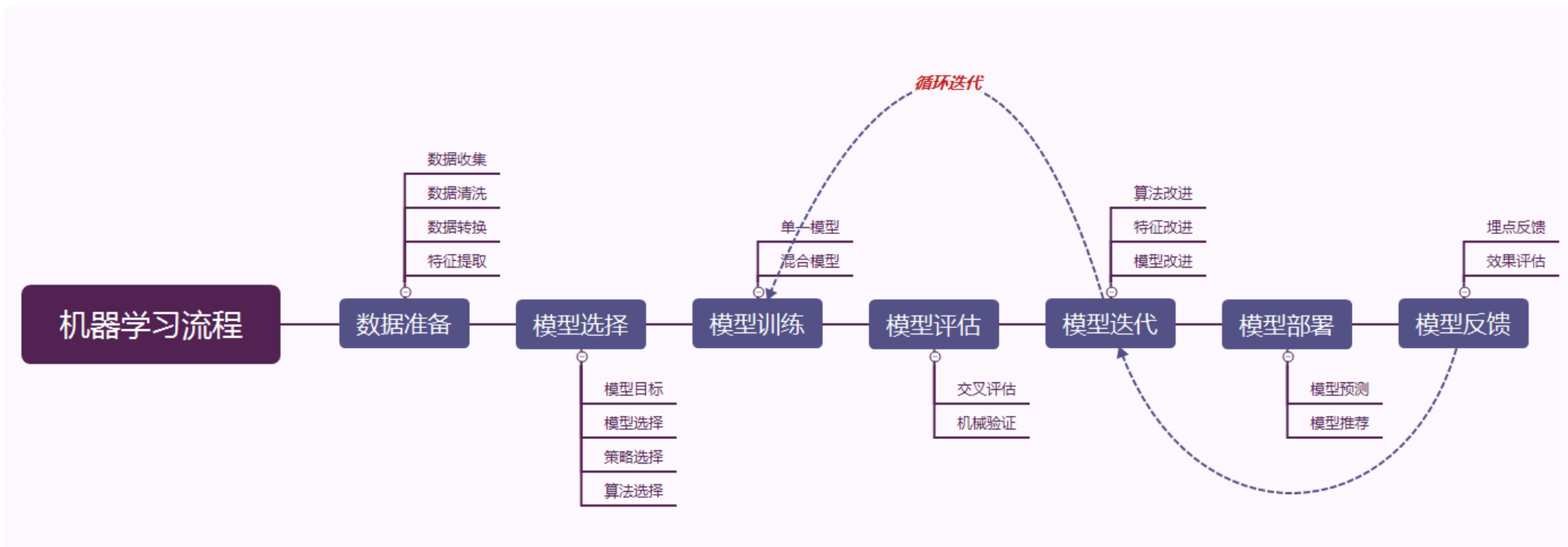
- 模型部署
 - 模型序列化
-
- Q&A：现有案例业务讨论

4.6 模型处理：案例练习

- 电影数据MovieLens
 - 建模：用户+电影喜好
 - 分类1：判断用户对某个电影的喜欢程度
 - 回归2：判断用户对某个电影的评分
-
- 案例：jupyter
 - Q&A：现有案例业务讨论

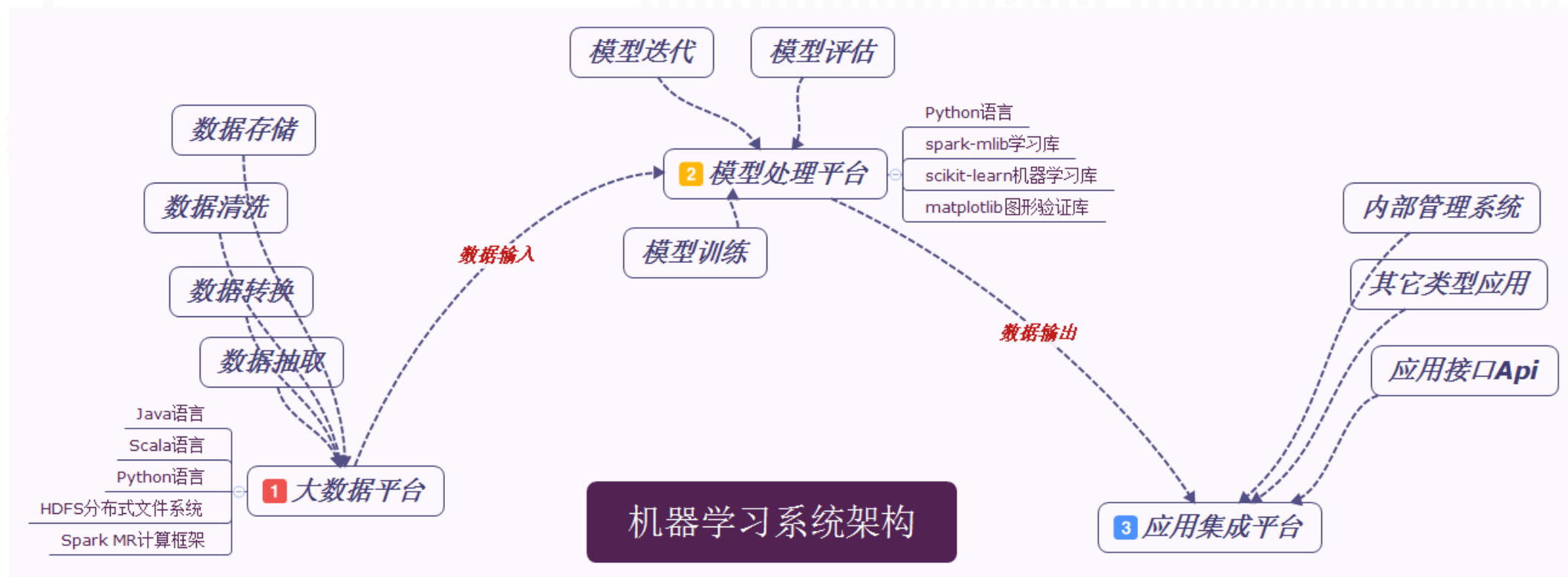
5.1 总结：机器学习流程

- 机器学习流程示意图
- Q&A：现有案例业务讨论



5.2总结：机器学习系统架构

- 机器学习系统架构示意图
- Q&A：现有案例业务讨论



5.2 总结：机器学习团队

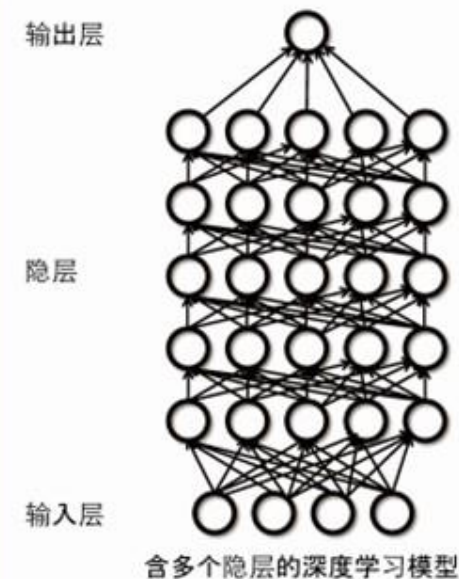
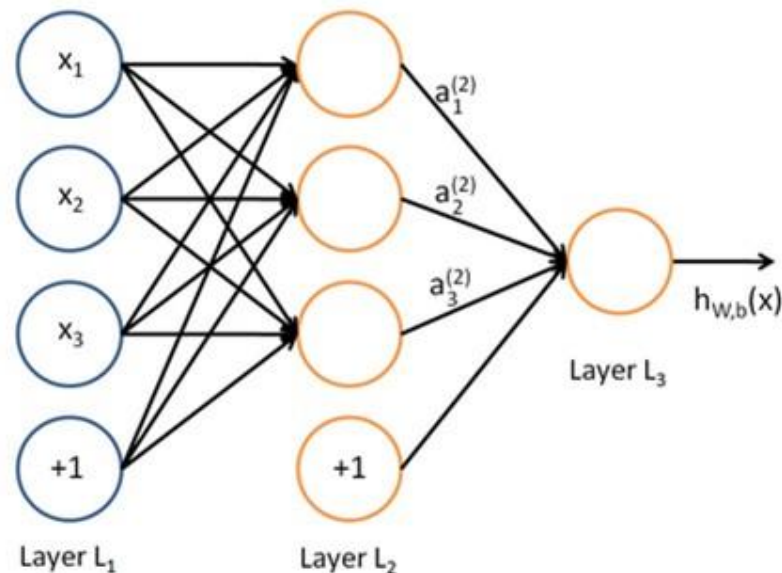
- 机器学习团队组织架构问题
- 团队分组：数据处理、模型处理
- 分工明确：模型理论/实战反馈
- Q&A：现有案例业务讨论

5.3 总结：机器学习思维

- 理论派：深入了解数学模型本质；
 - 实战派：熟练了解不同模型的背景、优越性、局限
 - 混合派：跨语言平台能力
-
- Q&A：现有案例业务讨论

5.4总结：机器学习发展

- 深度学习概念
- 浅层学习概念
- 示例图：神经网络
- Q&A：现有案例业务讨论





“给我一个支点，我就能撬起整个地球。”

▲ 阿基米德

联系方式

- Mail:1789909854@qq.com
- QQ:1789909854
- [Tel:17621063575](tel:17621063575)
- Wechat:ynuosoftware

