预备知识

徐杨

作为一名数据分析师，
编程水平、业务知识、行业经验等 决定了职业发展的下限。

而数学，往往决定了职业发展的上限。

让我们用几天的时间，为今后数据分析的职业发展开个好头。

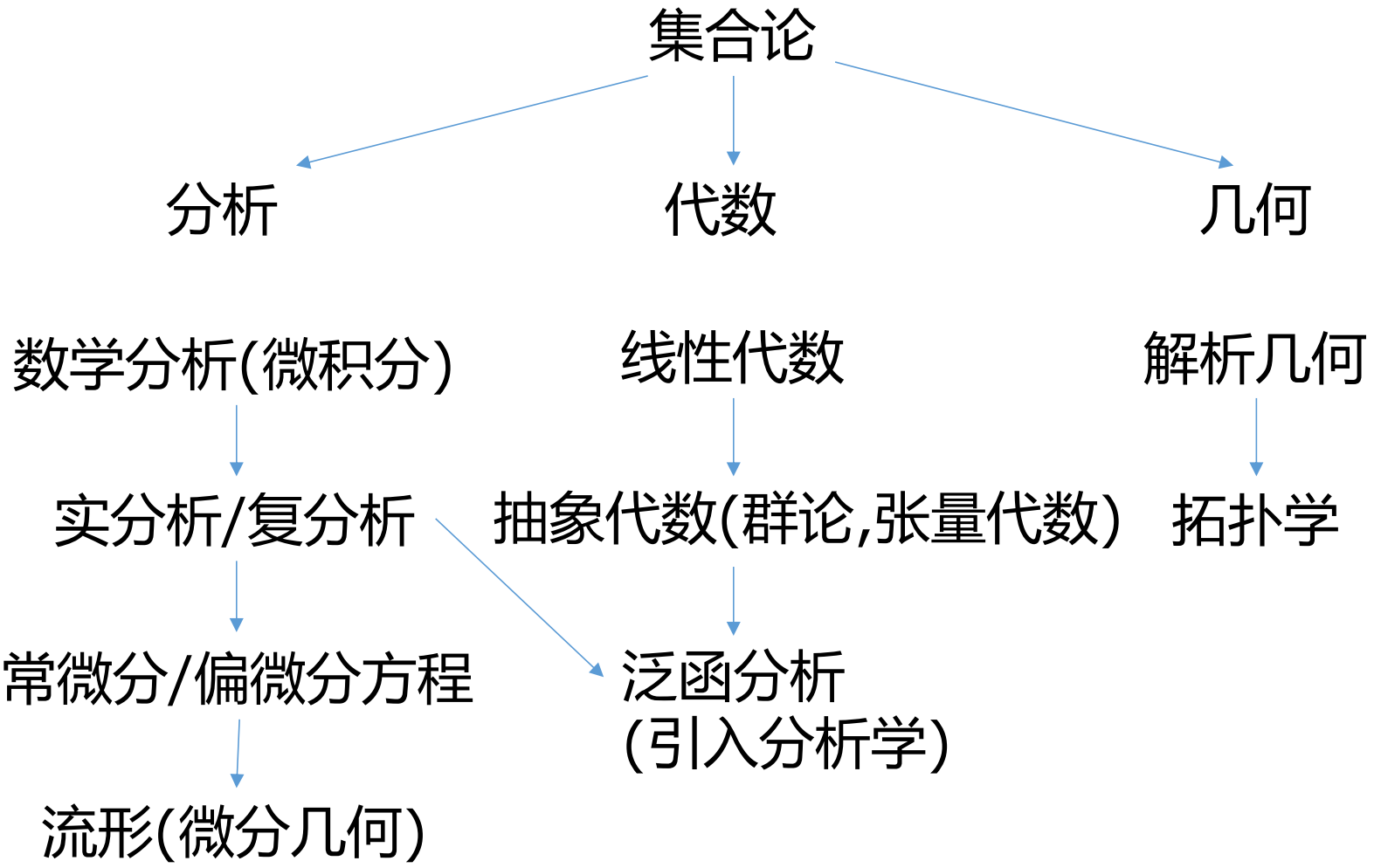
- **课程共四天，包含内容如下**

- 第一天：预备知识，线性代数
- 第二天：函数，微积分
- 第三天：数据度量，统计量及抽样分布，参数估计
- 第四天：假设检验，相关分析，回归分析

第一节

数学概况

数学概况


集合论
分析
代数
几何
数学分析(微积分)
线性代数
解析几何
实分析/复分析
抽象代数(群论,张量代数)
拓扑学
常微分/偏微分方程
泛函分析
(引入分析学)
流形(微分几何)

第二节

数据类型

数据类型

- 离散型数据

离散随机变量是指一个只取有限个或可数无限个数值的随机变量。
通常用古典概型来描述。

- 连续型数据

连续随机变量是指一个取任何实数的概率都为零的变量。
通常用几何概型来描述。

数据类型

- 横截面数据
- 时间序列数据
- 面板数据

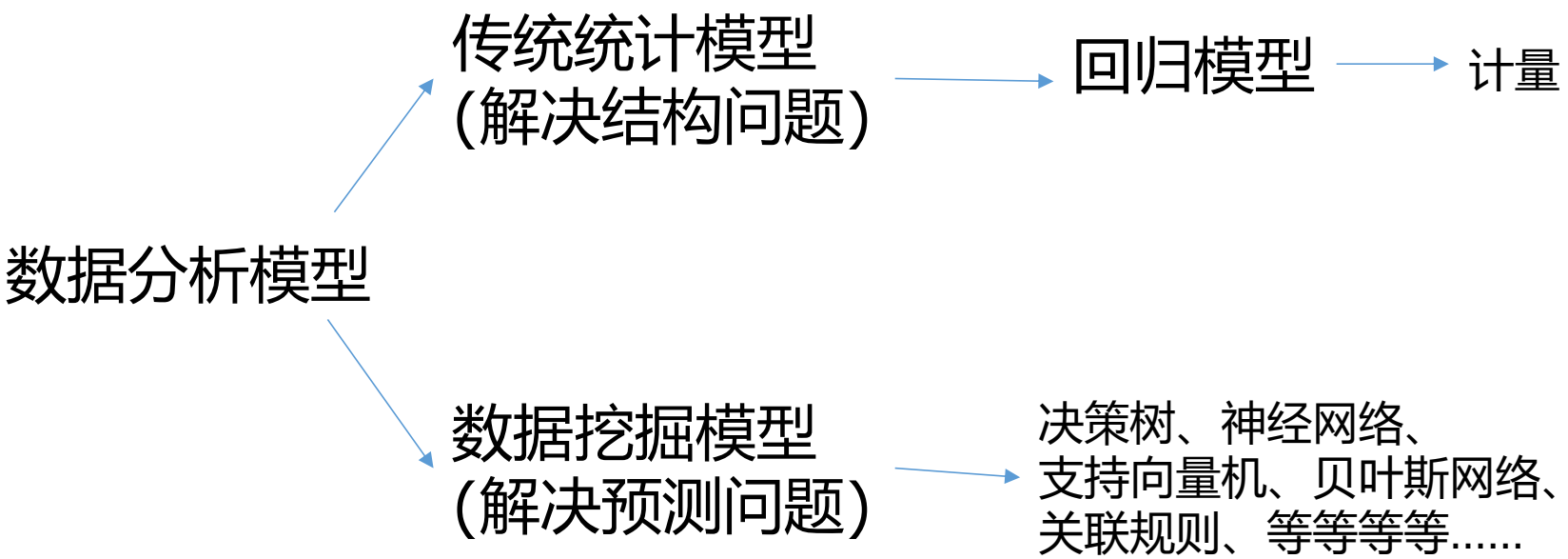
数据类型

排序	计算	数据类型	例
No	No	定类型	国籍
Yes	No	定序型	健康状况
Yes	Yes	数值型	时间

第三节

数学模型简介

数学模型


数据分析模型
传统统计模型
(解决结构问题)
回归模型
计量
数据挖掘模型
(解决预测问题)
决策树、神经网络、
支持向量机、贝叶斯网络、
关联规则、等等等等……

回归模型

最简单的回归模型是一元线性回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

举些例子来理解模型：

- LOL胜率 $= \beta_0 + \beta_1$练习时间 $+ \epsilon$

  练习时间每多一分钟，会使胜率提高β_1个百分点。

- 皮肤光泽度 $= \beta_0 + \beta_1$燕窝摄入量 $+ \epsilon$

  每多吃一斤燕窝，会使皮肤光泽提高β_1度。

回归模型

最简单的回归模型是一元线性回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

将其扩展到多个自变量的形式，即为多元线性回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

将其扩展到非线性形式形式，即为广义回归模型：

$$Y = f(X_1, X_2, \cdots, X_p) + \varepsilon$$

回归模型

横截面模型：
多元回归、逻辑回归、托宾回归、截尾回归

时间序列模型：
ARIMA、GARCH、协整

面板数据模型：
固定效应/随机效应、空间计量模型

结构方程模型

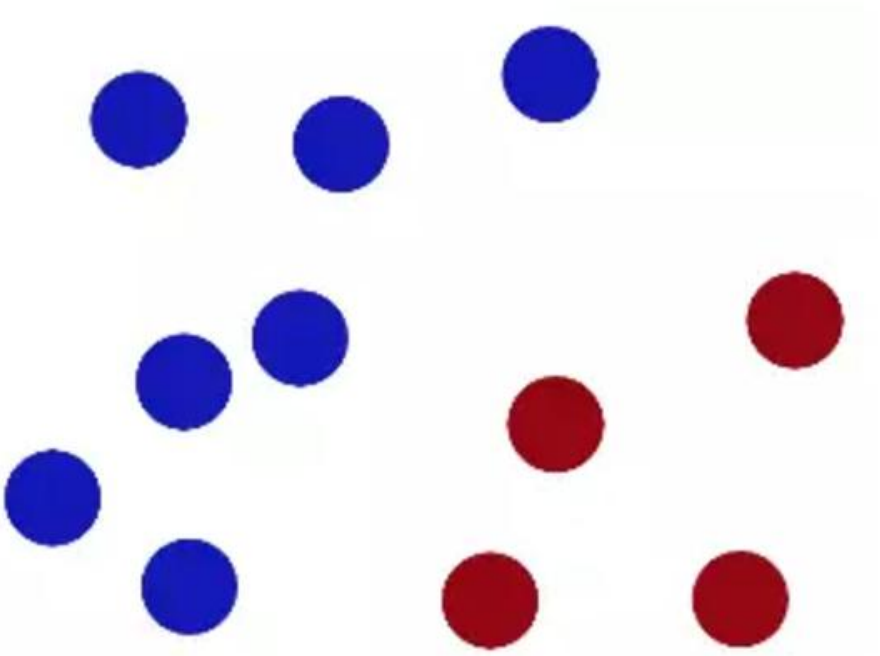
向量自回归模型（VAR）

支持向量机

私人干货之

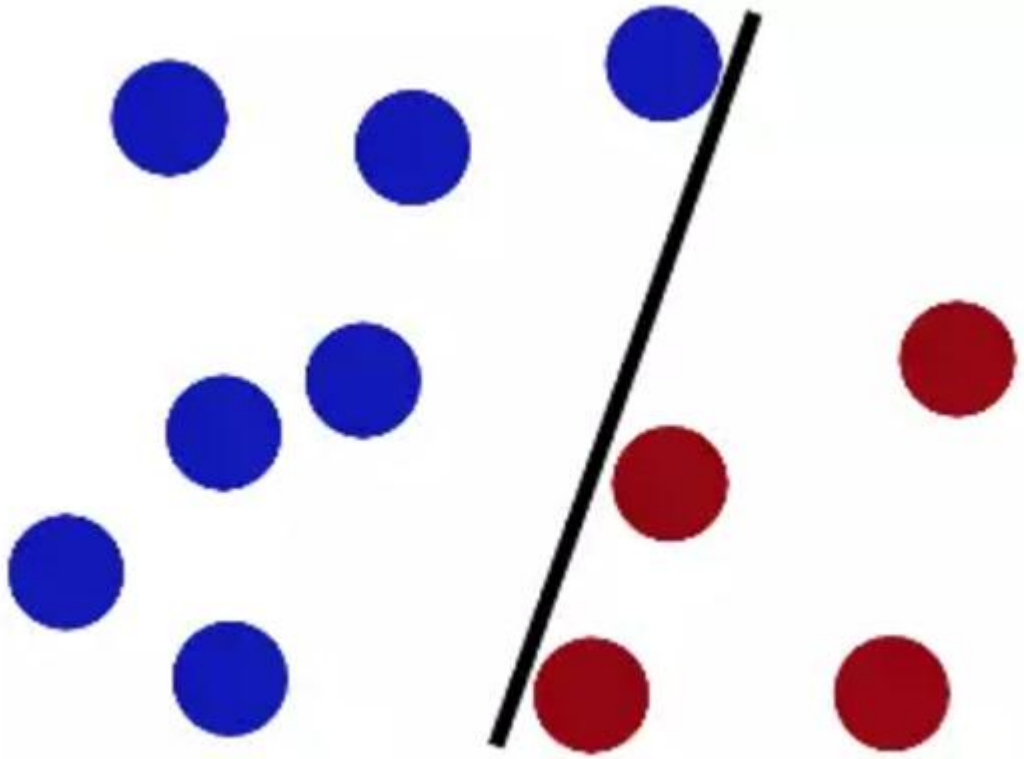
支持向量机

支持向量机



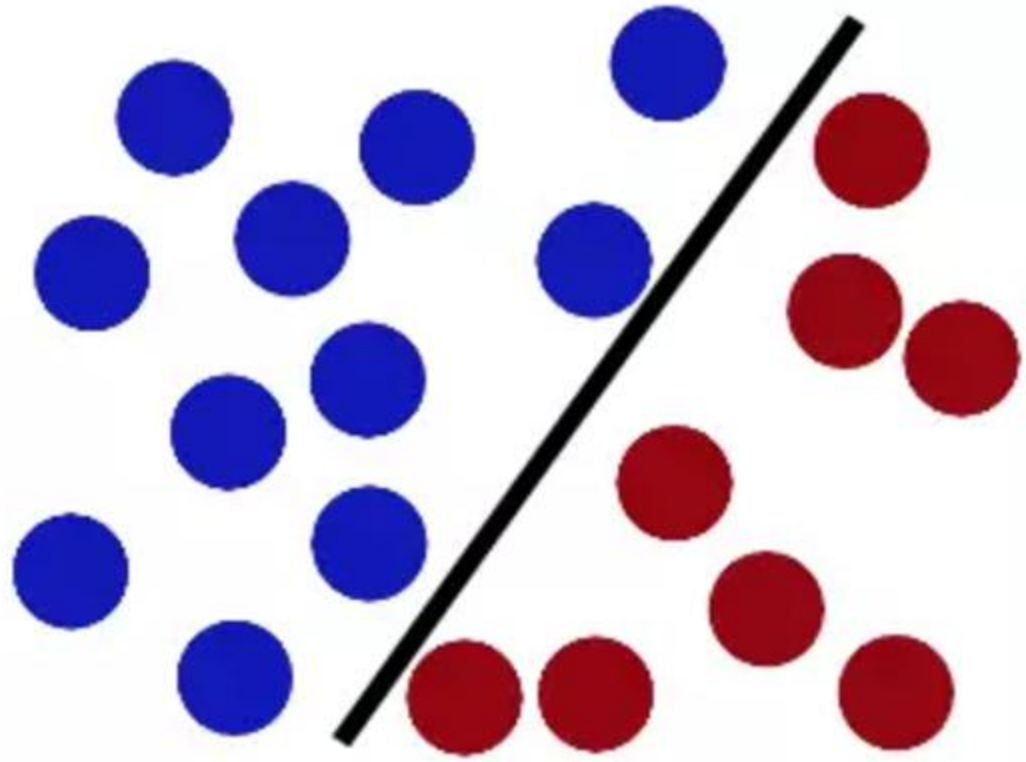
请用一条直线将2种类型的球分开

支持向量机



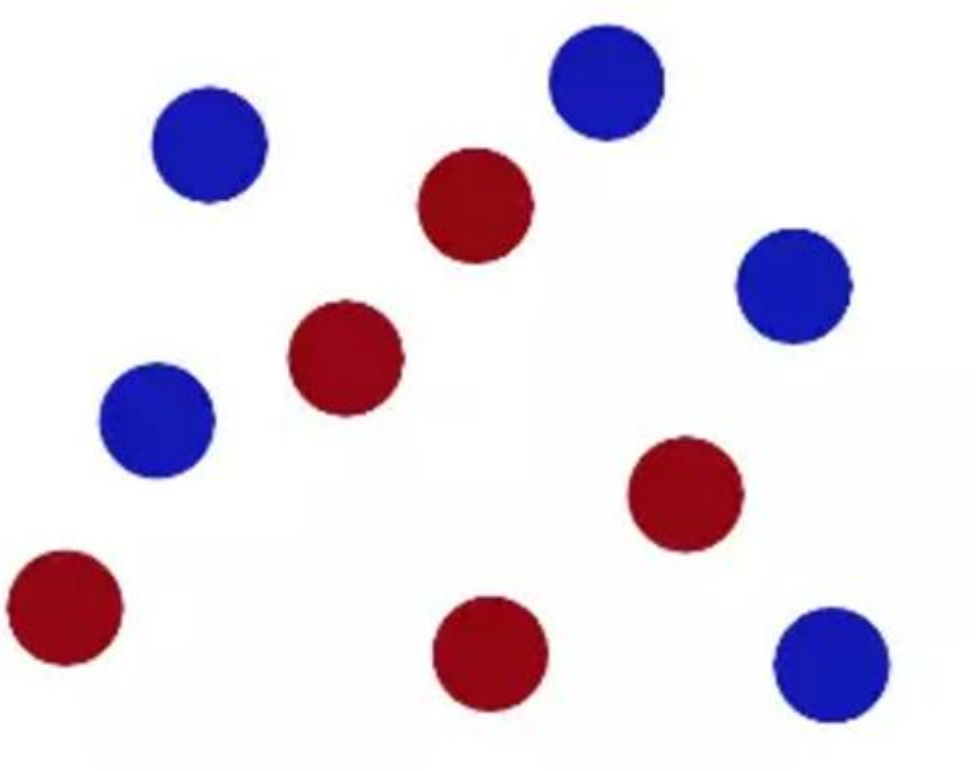
线性分类

支持向量机



线性分类

支持向量机



但是非线性分类怎么办?

支持向量机