



北京邮电大学

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS



Data Analytics with R
语言数据分析



所谓学习，归类而已

艾新波 / 2018 • 北京



课程体系

R语言数据分析

上部：论道



- 第1章 气象万千、数以等观
- 第2章 所谓学习、归类而已
- 第3章 格言联璧话学习
- 第4章 源于数学、归于工程

中部：执具



- 第5章 工欲善其事必先利其器
- 第6章 基础编程
- 第7章 数据对象

- 第8章 人人都爱tidyverse

- 第9章 最美不过数据框

下部 博术



- 第10章 观数以形
- 第11章 相随相伴、谓之关联
- 第12章 既是世间法、自当有分别
- 第13章 方以类聚、物以群分
- 第14章 庐山烟雨浙江潮

以机器学习为内核

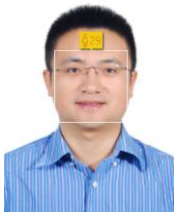
数据分析 \approx 机器学习 / 数据挖掘

\approx 认识数据 + 关联 + 分类 + 聚类

\approx 寻找关系结构 (核心是归归类)

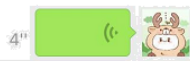
数据分析：发现数据背后的规律

年龄识别



<https://www.how-old.net/>

语音转文字

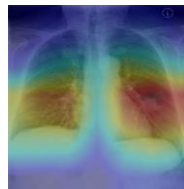


现在开始语音转文字测试。

转换完成

发现数据
背后的规律

疾病诊断



个性化推荐



猜你喜欢



每日推荐



音乐圈

规律是什么？

模式是什么？

数学公式？

.....



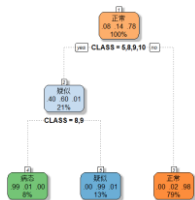
规律：某种关系结构

- ❑ 从系统理论的角度看，世界由系统组成，一切客观存在均可视为系统
- ❑ 系统又由什么组成呢？元素及其关系结构。所谓的规律，显然不属于元素的范畴，而是属于关系结构的范畴
- ❑ 《工程技术中的现代数学》：序关系、运算关系和映射是最基本的关系。各种复杂的关系形成各种结构。具有某种关系结构的数学对象集合就是某个特定的抽象数学空间，是现代数学的基本研究对象
- ❑ 机器所能学习的规律/模型，或者说数据中能挖出的规律，其实质都是：
某种关系结构

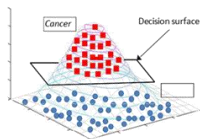
机器所能学到的：某种关系结构

主题分类	算法	关系（规律）
相关性分析	Pearson相关系数、Kendall相关系数等	相关关系
有监督的学习 (分类、回归)	决策树、逻辑斯蒂回归、朴素贝叶斯、支持向量机、人工神经网络等	因变量与自变量之间的映射关系，或者说函数关系
无监督学习 (聚类分析)	k-Means、层次聚类、DBSCAN等	数据空间中点与点、类与类之间的距离关系
关联规则	Apriori、FP-Growth等	项集之间的伴随关系
.....

有监督学习：映射关系



一棵树

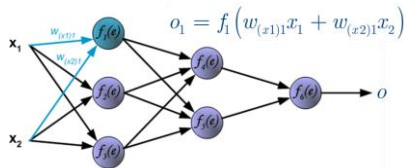


一个空间分隔超平面

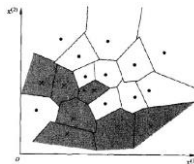
$$y = f(x)$$

$$p(y|x) = \frac{1}{Z(x)} \prod_{t=1}^T \Psi_t(y_t, y_{t-1}, x_t)$$

若干概率组合

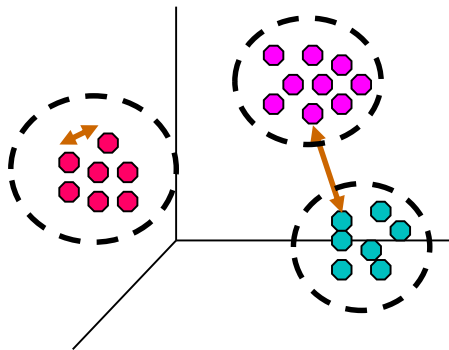


一个计算节点网络



基于距离的空间划分

无监督学习：距离关系



特征形成数据空间，每条观测记录都是该空间的一个点，聚类分析就是根据点之间的距离，把每组离得近的一些点划归为一类，聚类结果是类和它的特征或描述

一切都是关系结构

机器所能学到的，无非是某种：

关系结构

尤其是

变量之间的**关系**和数据空间的**结构**



一切都是关系结构

机器学习

存在的全部意义

只是为了寻找或表达

某种**关系结构**



A decorative blue border with rounded corners frames the entire slide. Two thin blue crosshair symbols are positioned on the right and left sides of the text.

谢谢聆听
Thank you

教师个人联系方式

艾新波

手机: 13641159546

QQ: 23127789

微信: 13641159546

E-mail: 13641159546@126.com

axb@bupt.edu.cn

地址: 北京邮电大学科研楼917室

课程网址: <https://github.com/byaxb/RDataAnalytics>

