

Machine Learning and Pattern Recognition

东北大学 “智能+X” 新工科课程系列

# 机器学习与模式识别

东北大学 信息科学与工程学院  
人工智能系、智能感知与机器人研究所  
陈东岳



Basic Tasks of Pattern Recognition

# 模式识别基本任务

---



## CHAPTER ONE

# 从IRIS数据集开始

Start from IRIS Data Set



# 一、从IRIS数据集开始

## ▶ 数据集来源

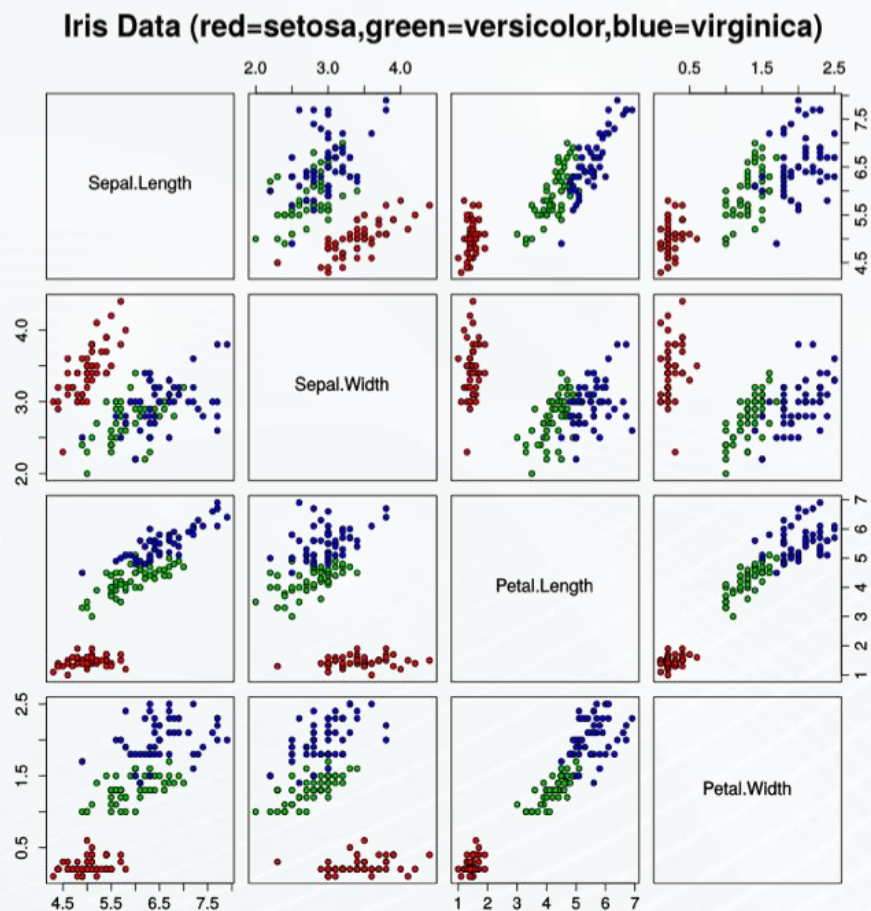
- 1935年，美国植物学家Edgar Anderson，采集三种鸢尾花形态学数据；
- 1936年，英国统计学家Ronald Fisher，整理出Fisher's Iris数据集；

## ▶ 数据集介绍

- 3个子类：Setosa, Versicolor, Virginica
- 每个子类50株花卉，共150株；
- 每朵花5个数据：
  - 花萼宽度、花萼长度、花瓣宽度、花瓣长度、花卉种类

# 一、从IRIS数据集开始

## ▶ 数据分布



## CHAPTER TWO

# 基本任务

Basis Tasks



## 二、基本任务

### ► 分类——Classification

- 定义

- 根据观测数据判断一个对象的类别

- IRIS数据库示例

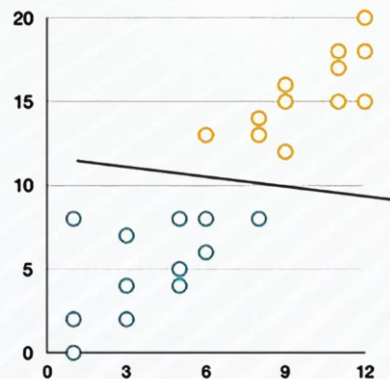
- 根据鸢尾花的花萼长度、宽度与花瓣长度、宽度判断其类别

- 其他示例

- 依靠视觉判断衣服那边是正面;
  - 依靠嗅觉判断衣服是干净的还是穿过的;
  - 依靠味觉判断荷包蛋上放的是盐还是糖;
  - 依靠听觉判断手机铃声是否是自己的;
  - 依靠触觉从书包中摸出了手机;

- 典型应用

- 人脸识别, 无人驾驶, 医疗辅助诊断, 垃圾邮件过滤.....





## 二、基本任务

### ► 聚类——Clustering

- 定义

- 将观测数据分为由相似对象组成的多个类别

- IRIS数据库示例

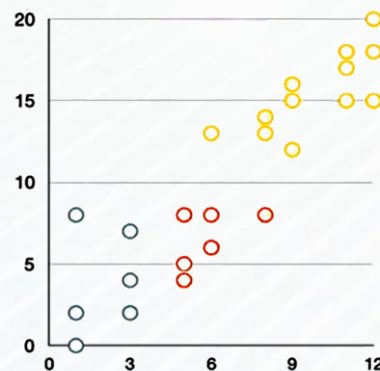
- 不考虑类别信息，将150朵鸢尾花根据形态学特征分为3类。

- 其他示例

- 将手机APP分为n个文件夹存储；
  - 将同品种的苹果分成3堆，不同定价；
  - 早期动物学家划分动物门类；
  - 保险推销员对客户进行分组；
  - 手机通讯录分组；

- 典型应用

- 互联网客户分析、图像分割、网站资源管理、故障诊断.....





## 二、基本任务

### ► 异常检测——Anomaly detection

- 定义

- 找出不属于某个类别的异常对象;

- IRIS数据库示例

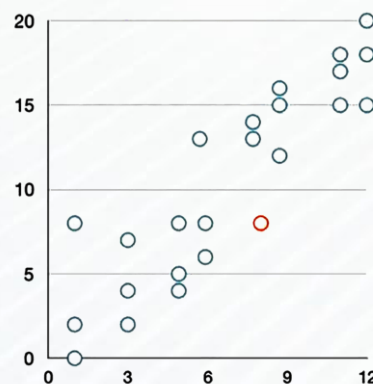
- 判断一朵新出现的鸢尾花是否属于已知的3类。

- 其他示例

- 根据照片判断是否认识某人;
- 在羊群中找出牧羊犬;
- 判断一个人是否生病了;
- 判断生产线是否正常运行;
- 判断电脑是否中了病毒;

- 典型应用

- 信用卡异常使用、账户异常登录、智能视频监控.....



## 二、基本任务

### ► 回归——Regression

- 定义

- 根据观测数据去预测对象的某个连续属性的数值

- IRIS数据库示例

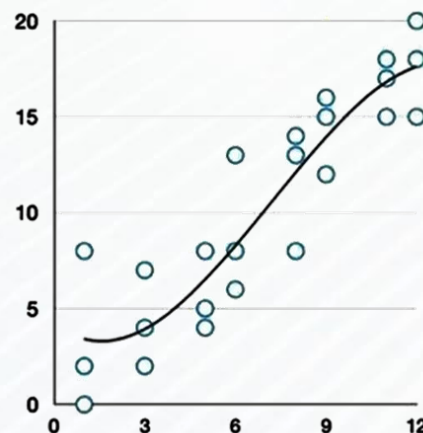
- 根据鸢尾花的花萼长度、宽度，预测其花瓣长度与宽度；

- 其他示例

- 预测明天的气温；
  - 给一辆二手车定价；
  - 估计一个同学的身高；
  - 预测三峡大坝今年的发电量；
  - 估计美国总统特朗普的选票；

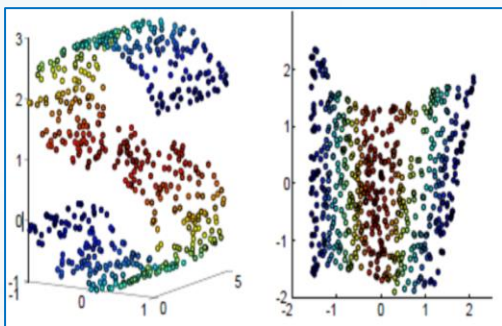
- 典型应用

- 天气预报、商品价格估计、目标视觉检测、医学检测.....



## 二、基本任务

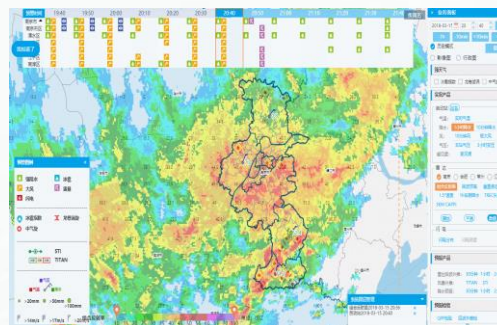
### ► 其他任务



降维



查询与匹配



概率估计



数据可视化



数据生成



数据结构化

# 例题



## 例题——判断模式识别任务类型

- 1. 根据鸢尾花的形态学数据估计其花期长短
  - 回归
- 2. 估计围棋比赛中某一时刻黑方的胜率
  - 概率估计
- 3. 无人驾驶汽车看到红灯时自动停车
  - 分类
- 4. 智能无人机群执行某任务时自动编队
  - 聚类
- 5. 从火星的航拍照片中找到非自然的地表形态
  - 异常检测
- 6. 博物馆对收藏品进行分组展览
  - 聚类
- 7. 百度搜图
  - 查询与匹配
- 8. 模糊人脸图像的超分辨率重建
  - 数据生成
- 9. 预测某只股票明天的涨跌情况
  - 分类

