

科技论文实验数据统计 分析与可视化

张立国



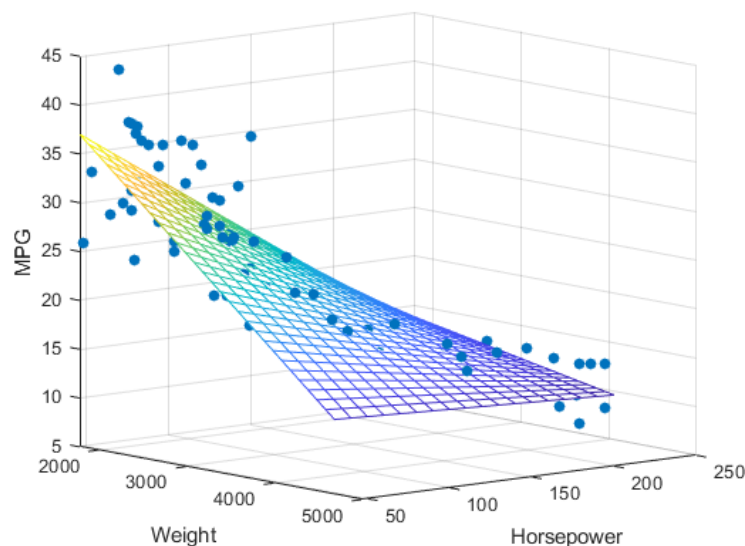
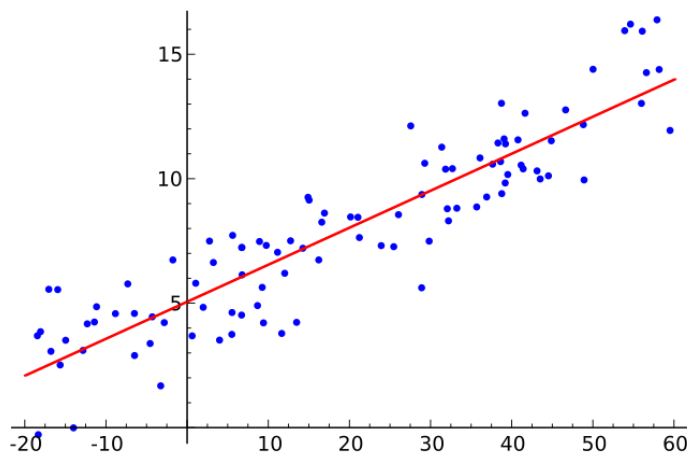
哈尔滨工程大学
Harbin Engineering University

大 工 至 善 · 大 学 至 真

数据分析

回归问题

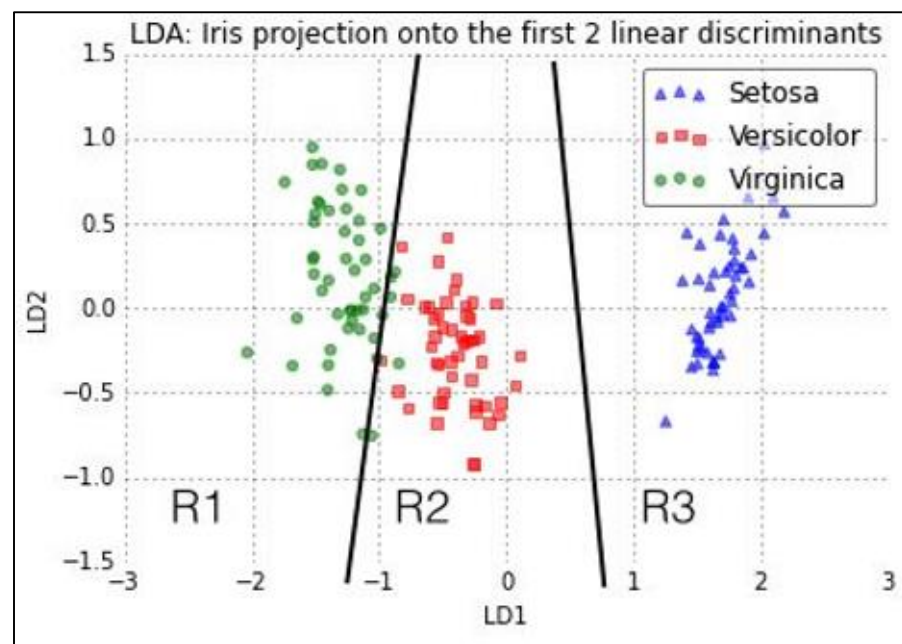
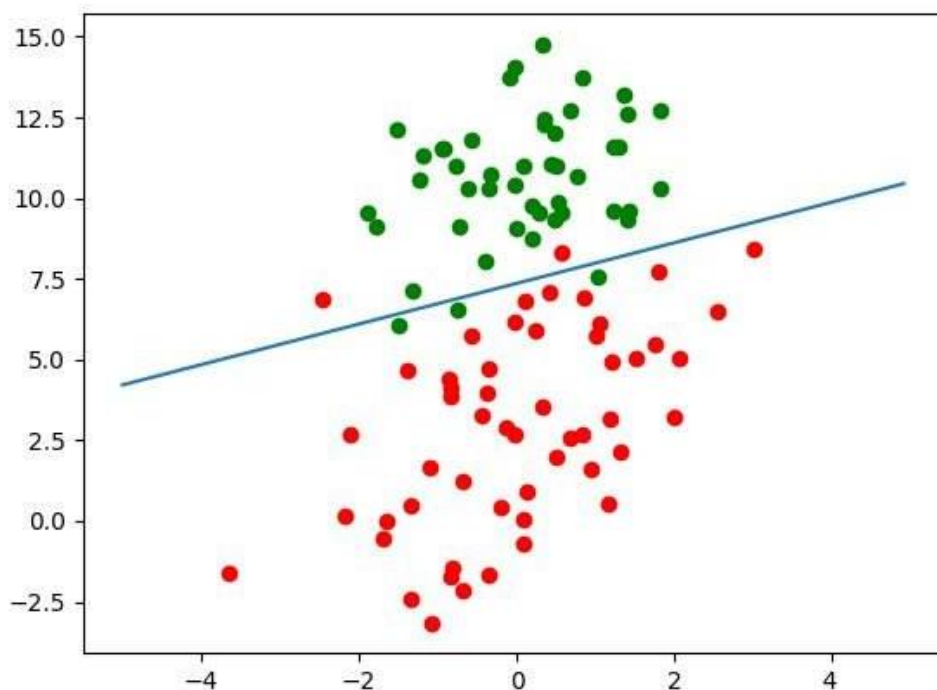
- 回归问题通常指**回归分析**，是一种统计学上分析数据的方法，目的在于了解两个或多个变量间是否相关、相关方向与强度，并建立数学模型以便观察特定变量来预测研究者感兴趣的变量。
- 回归分析是建立因变量 Y （或称依变量，反因变量）与自变量 X （或称独变量，解释变量）之间关系的模型。



数据分析

分类问题

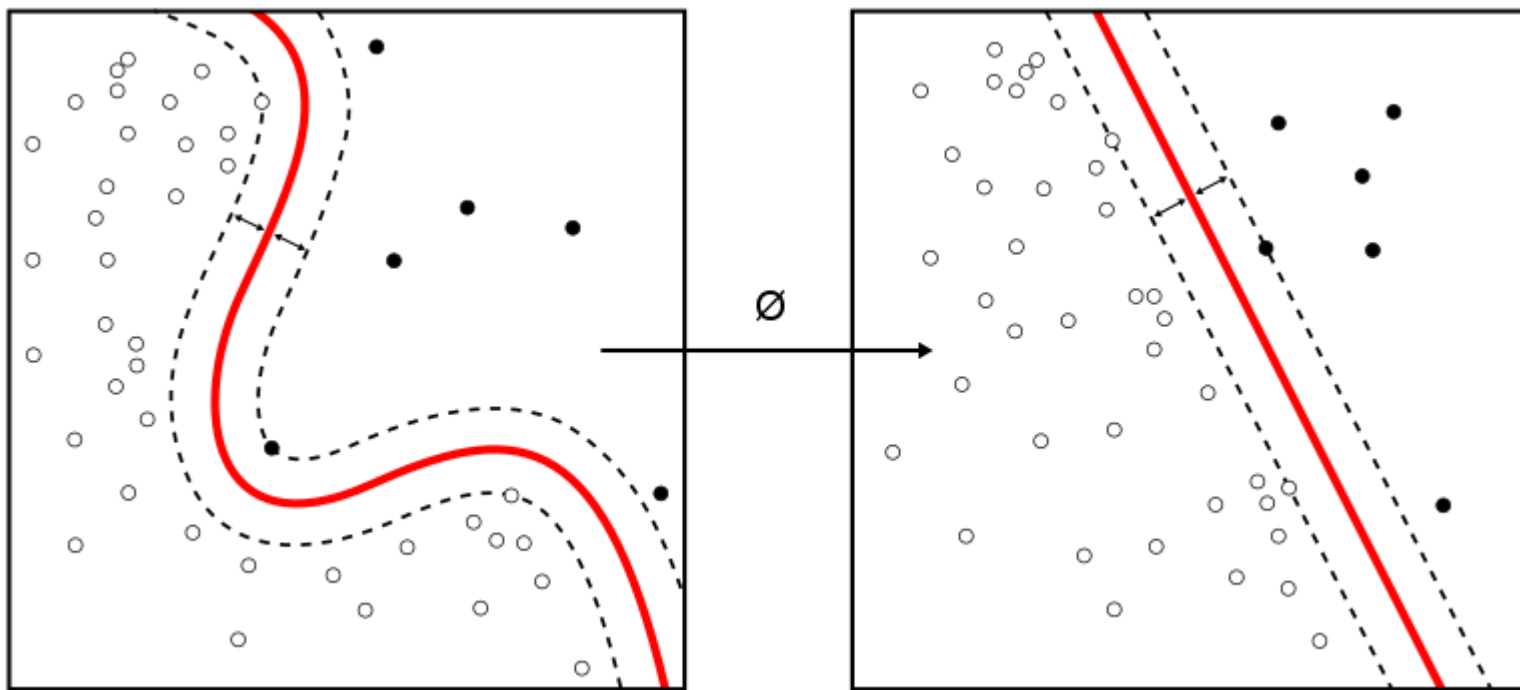
分类是根据某些**特征**，将“事物”区分并分配到不同的种类里。种类的数量决定了分类属于**二分类**问题还是**多分类**问题。



数据分析

分类问题

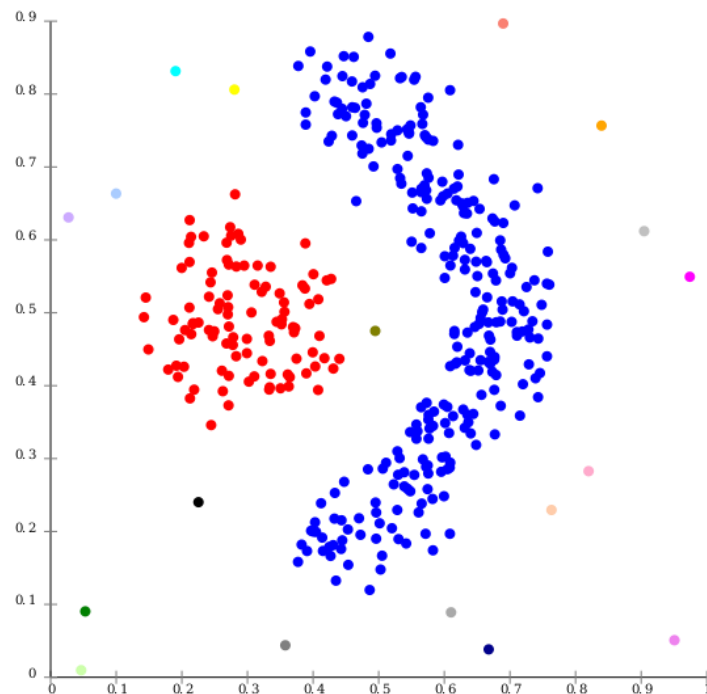
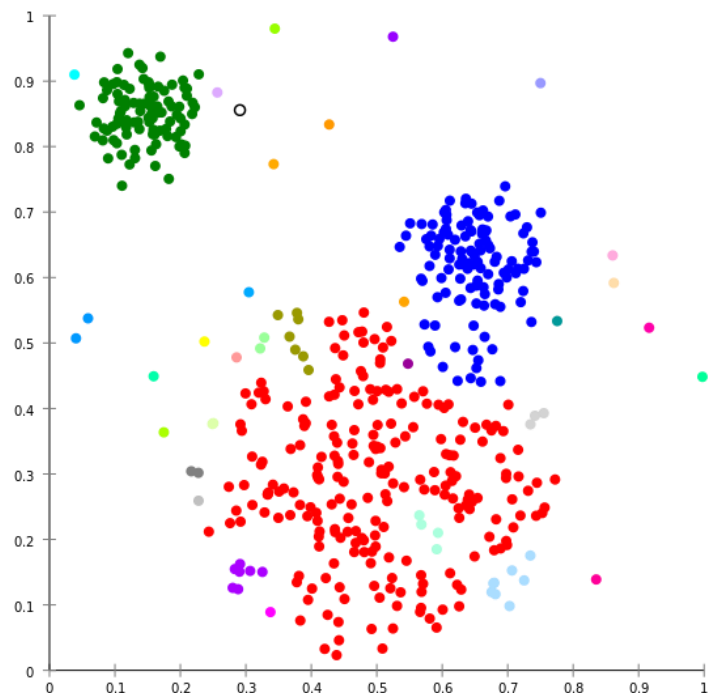
分类是根据某些**特征**，将“事物”区分并分配到不同的种类里。种类的数量决定了分类属于**二分类**问题还是**多分类**问题。



数据分析

聚类问题

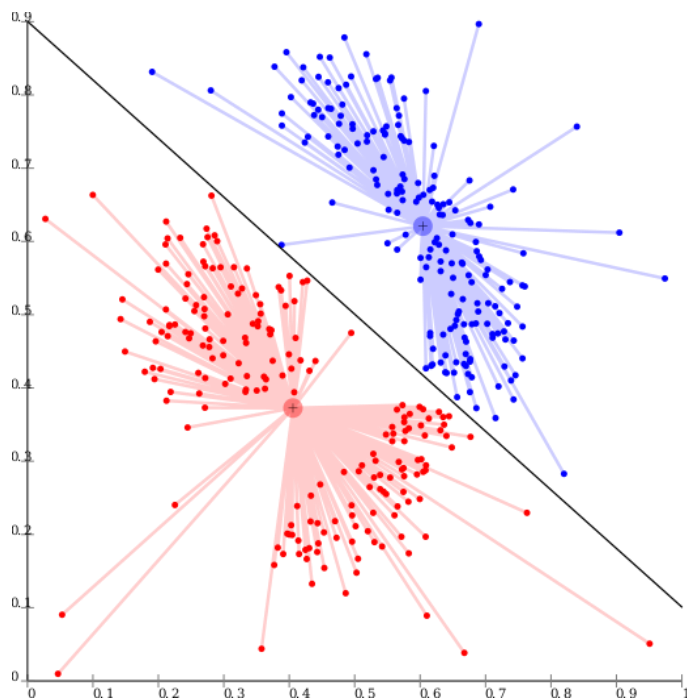
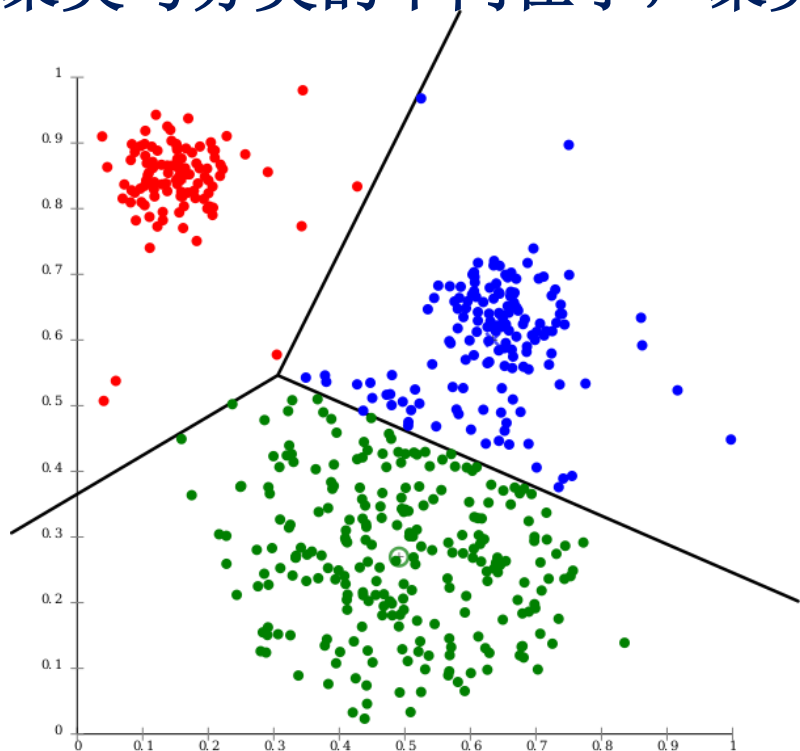
聚类问题是根据某些特征，试图将所给数据集中的样本划分为若干个互不相交的子集，每个子集称之为一个“簇”。聚类与分类的不同在于，聚类所要求划分的类是未知的。



数据分析

聚类问题

聚类问题是根据某些特征，试图将所给数据集中的样本划分为若干个互不相交的子集，每个子集称之为一个“簇”。聚类与分类的不同在于，聚类所要求划分的类是未知的。

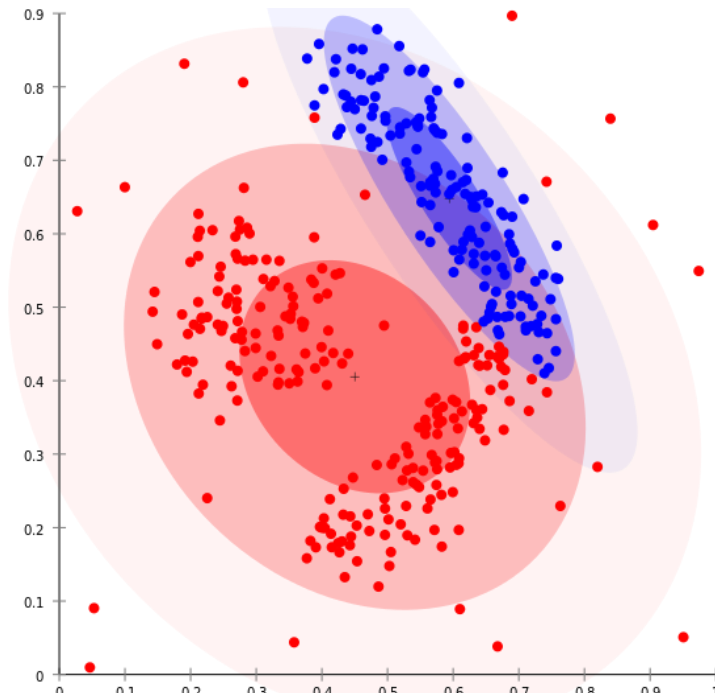
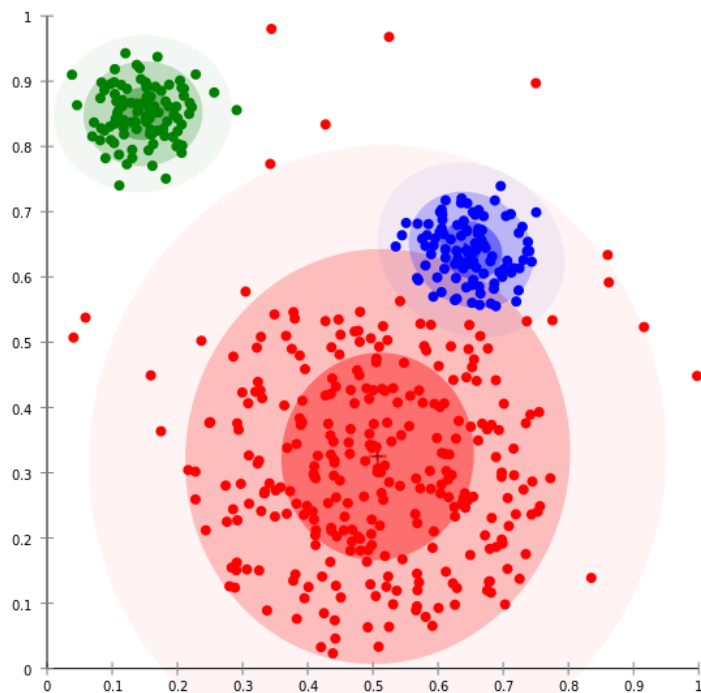


k -均值聚类

数据分析

聚类问题

聚类问题是根据某些特征，试图将所给数据集中的样本划分为若干个互不相交的子集，每个子集称之为一个“簇”。聚类与分类的不同在于，聚类所要求划分的类是未知的。



高斯混合模型聚类

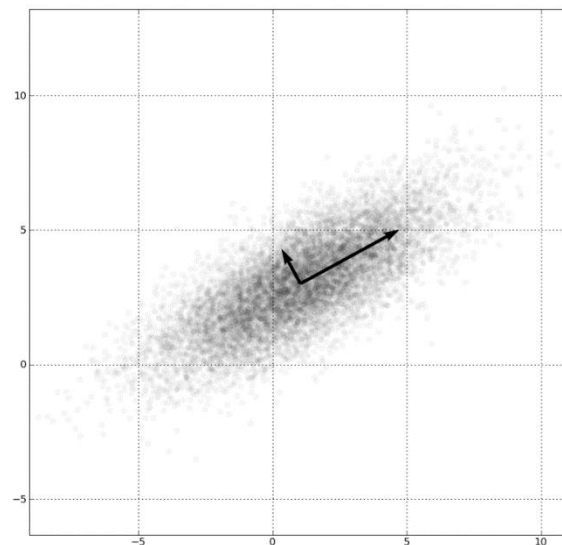
数据分析

主成分分析

主成分分析（**Principal components analysis, PCA**）是一种统计分析、简化数据集的方法。它利用正交变换来对一系列可能相关的变量的观测值进行线性变换，从而投影为一系列线性不相关变量的值，这些不相关变量称为主成分（**Principal Components**）。

主成分可以看做一个线性方程，其包含一系列线性系数来指示投影方向。**PCA**对原始数据的正则化或预处理敏感（相对缩放）。

一个高斯分布，平均值为(1, 3)，标准差在(0.878, 0.478)方向上为3、在其正交方向上为1的主成分分析。黑色的两个向量是此分布的协方差矩阵的特征向量，其长度为对应的特征值之平方根，并以分布的平均值为原点。



数据分析

主成分分析

主成分分析（**Principal components analysis, PCA**）是一种统计分析、简化数据集的方法。它利用正交变换来对一系列可能相关的变量的观测值进行线性变换，从而投影为一系列线性不相关变量的值，这些不相关变量称为主成分（**Principal Components**）。

主成分可以看做一个线性方程，其包含一系列线性系数来指示投影方向。**PCA**对原始数据的正则化或预处理敏感（相对缩放）。

