



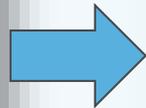
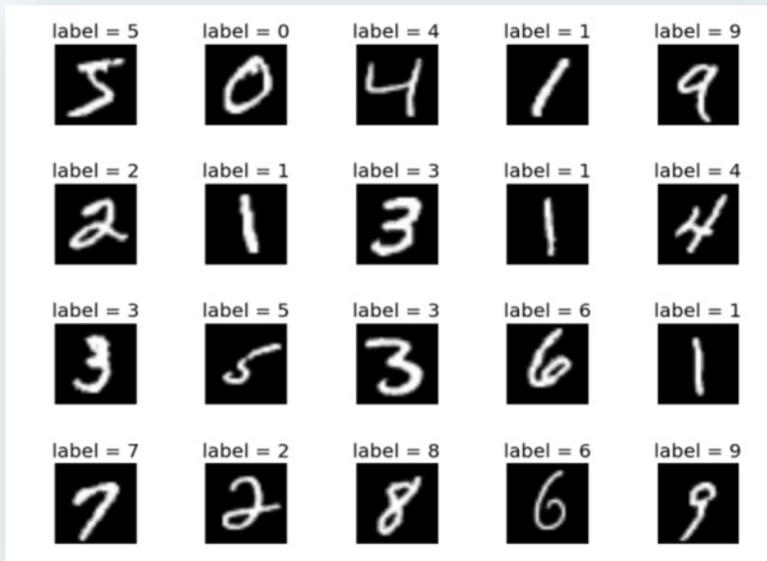
# 人工智能与信息社会

基于神经网络的智能系统II：监督学习和非监督学习

陈斌 北京大学 [gischen@pku.edu.cn](mailto:gischen@pku.edu.cn)

# 监督学习

- › 利用标注好信息的样本，经过训练得到一个模型，可以用来预测新的样本



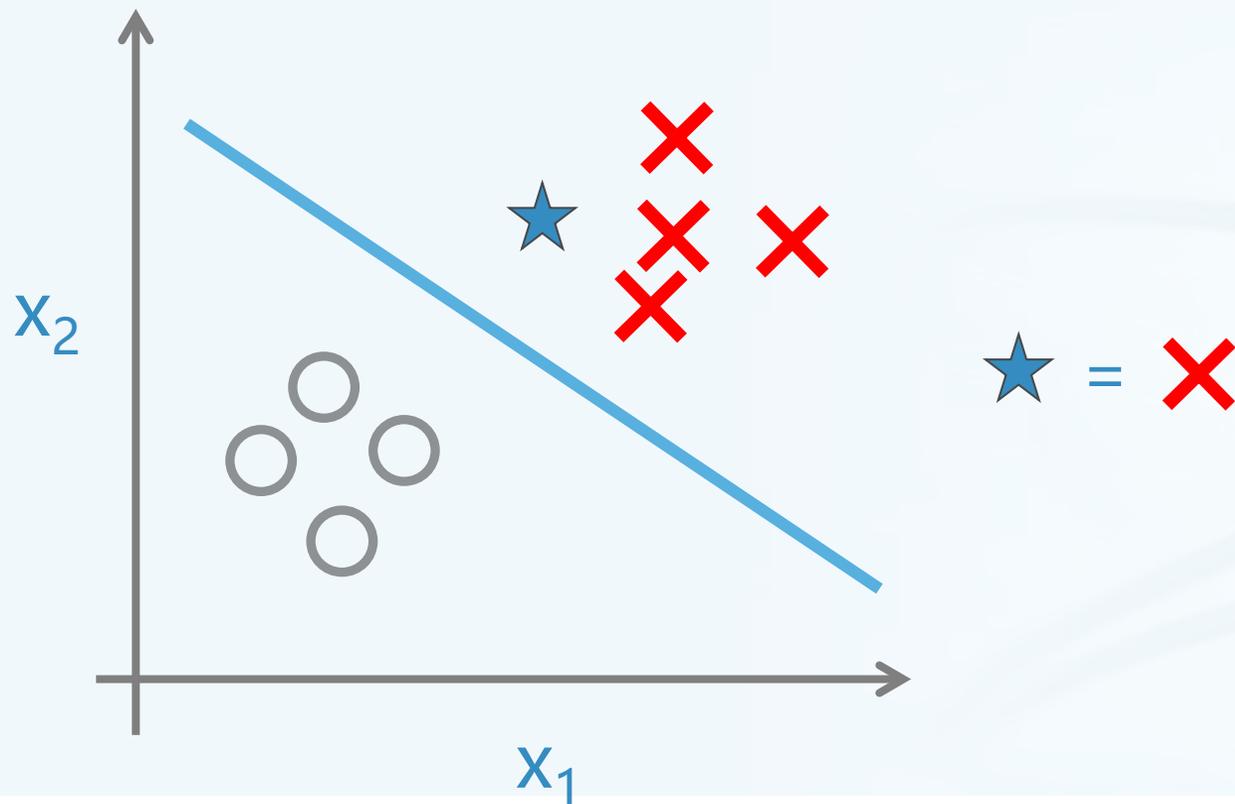
# 监督学习

- › 利用标注好信息的样本，经过训练得到一个模型，可以用来预测新的样本



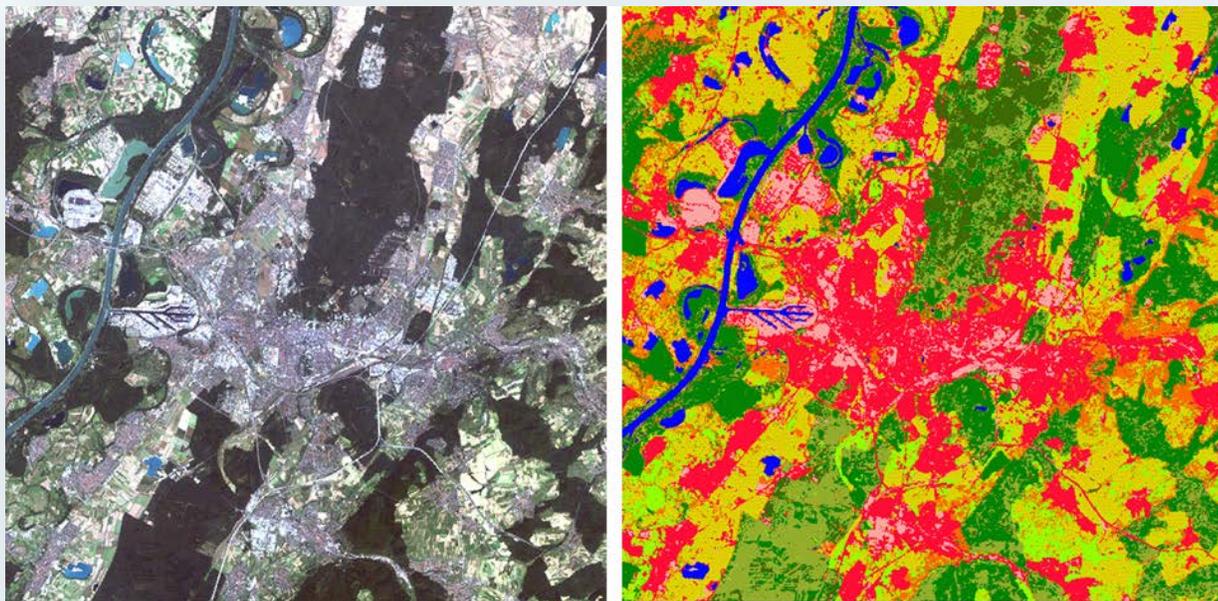
# 分类

› 当新来一个数据时，可以自动预测所属类型



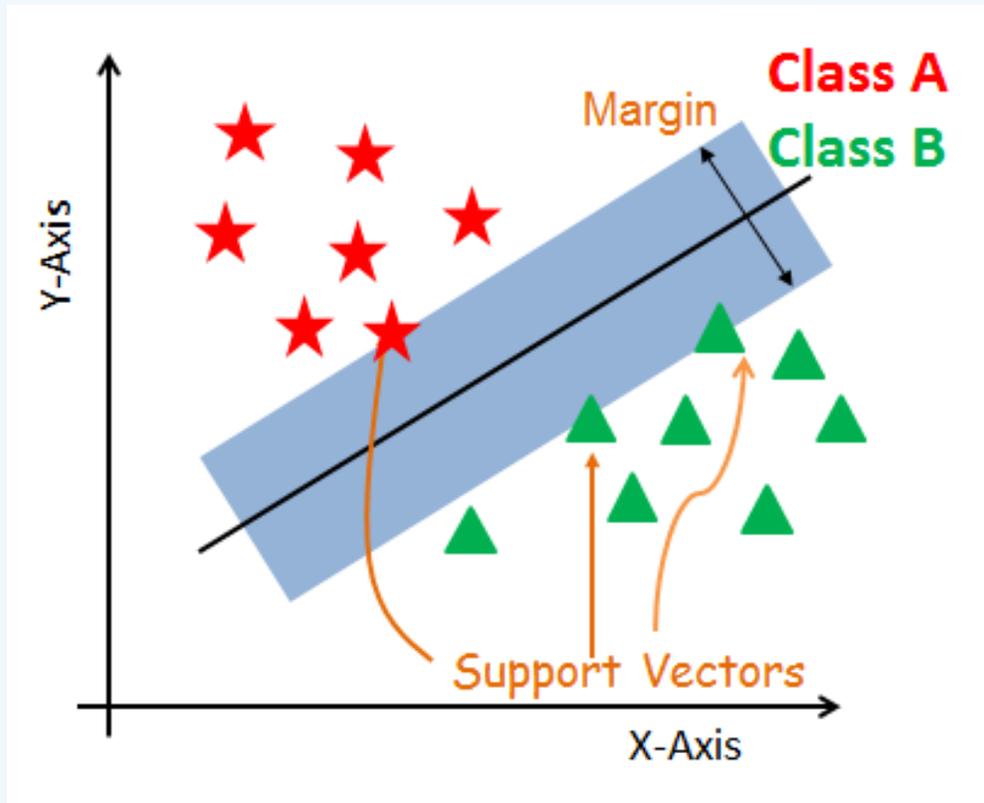
# 分类的应用

- 对于一幅遥感影像，对其中的部分水体、农田、建筑做好标记，通过监督分类方法就能得到其余的水体、农田、建筑。



# 分类（预测结果是离散值）相关的方法

## 支持向量机：寻找最大化样本间隔的边界



# 分类（预测结果是离散值）相关的方法

## 分类决策树

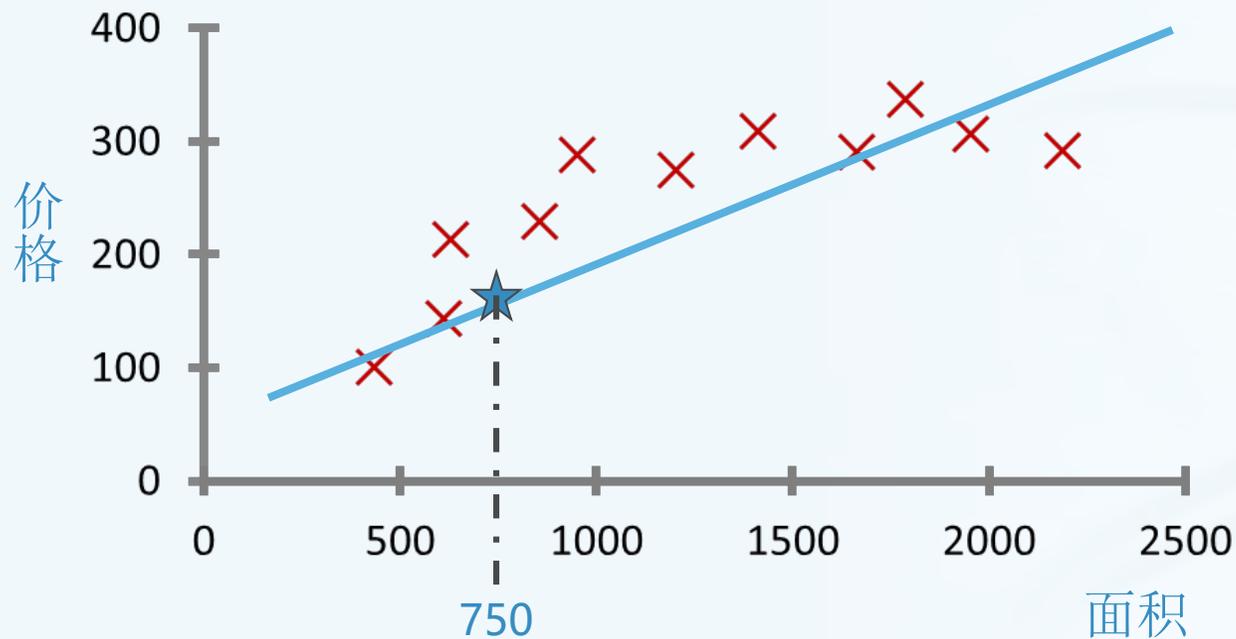




# 回归

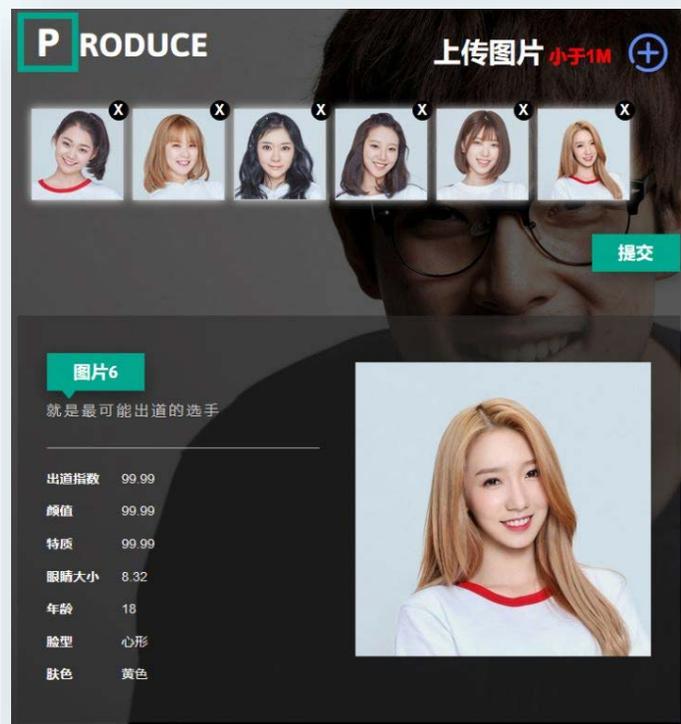
## 直线拟合 (最小二乘法)

我们希望通过已有的训练数据学习一个模型，当新来一个面积数据时，可以自动预测出销售价格



# 回归的应用

- 〉 人脸好看程度评分。通过标记分数的图片得出回归模型，输入新的图片就能得出分数。



# 回归（预测结果是连续值）相关的方法

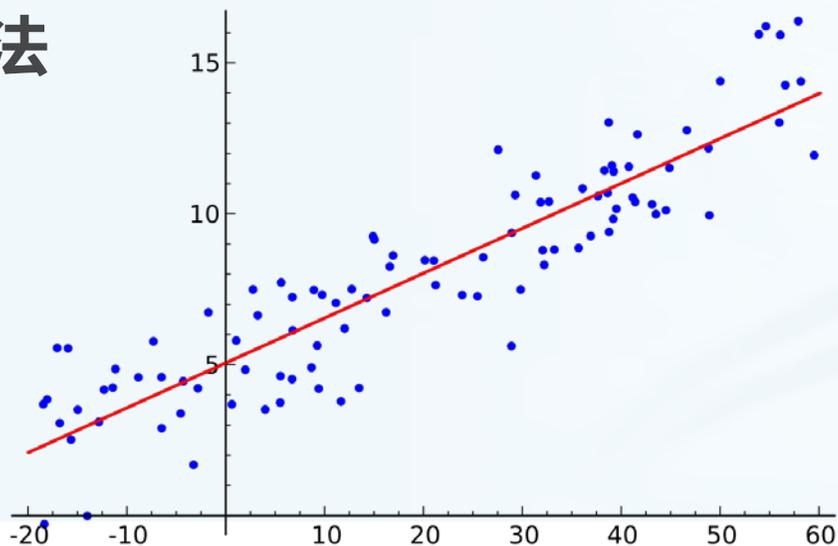
## › 线性回归

在平面上拟合线性函数

## › 最邻近方法

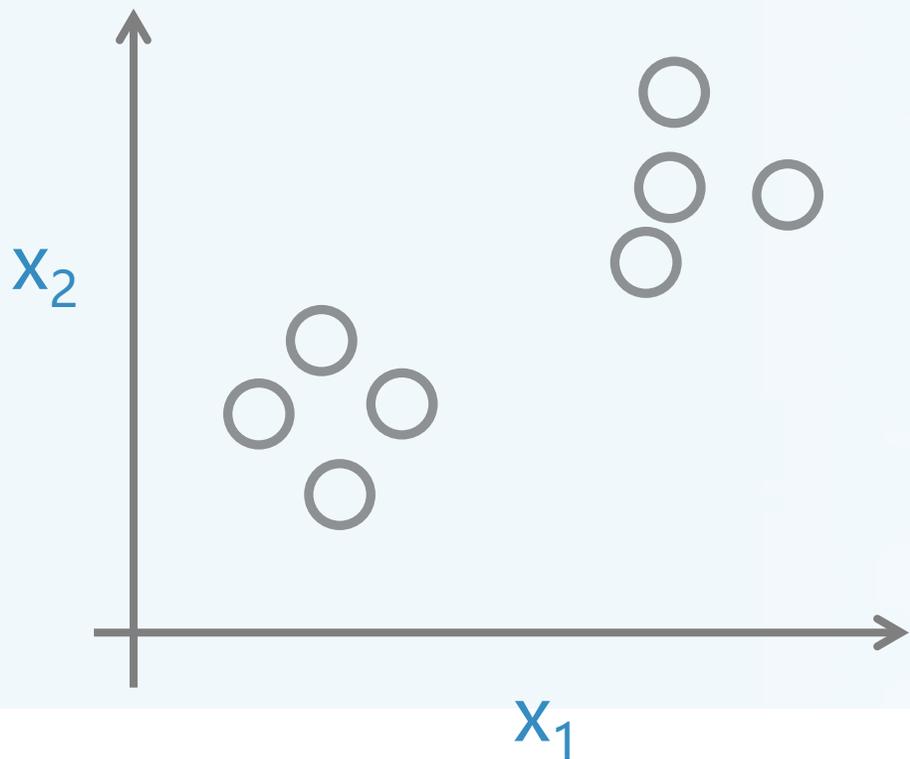
使用最相似的训练样本来预测新样本值

## › 神经网络方法



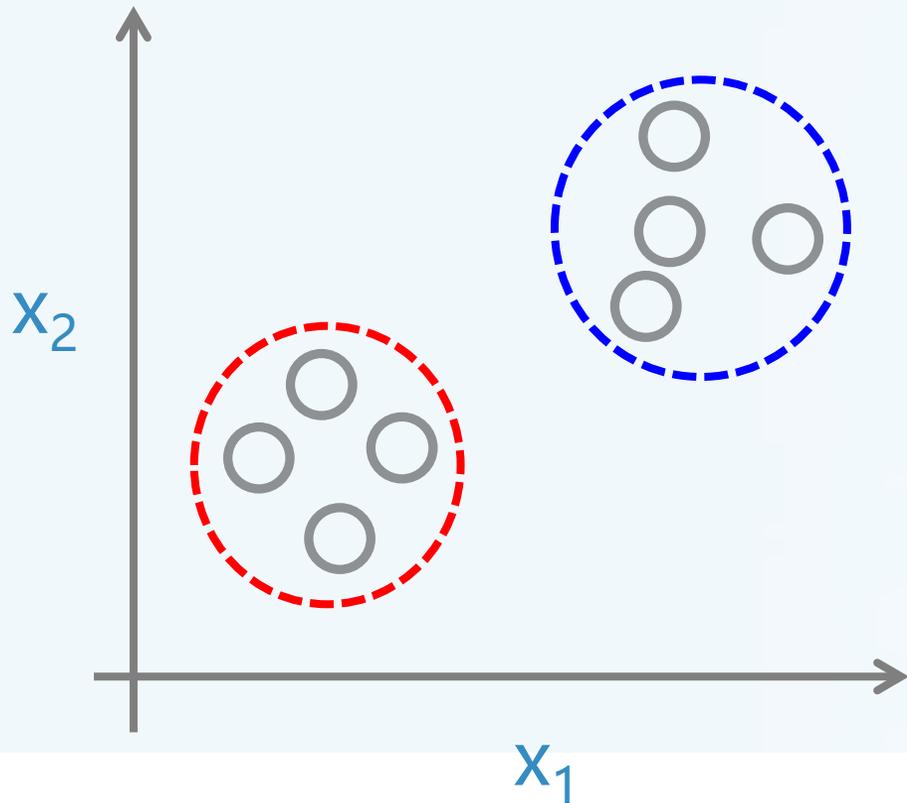
# 非监督学习

- 所有数据只有特征向量没有标签，但是可以发现这些数据呈现出聚群的结构，本质是一个相似的类型会聚集在一起。



# 聚类

- 把这些没有标签的数据分成一个一个组合，就是聚类（Clustering）。



# 聚类案例

## Google新闻

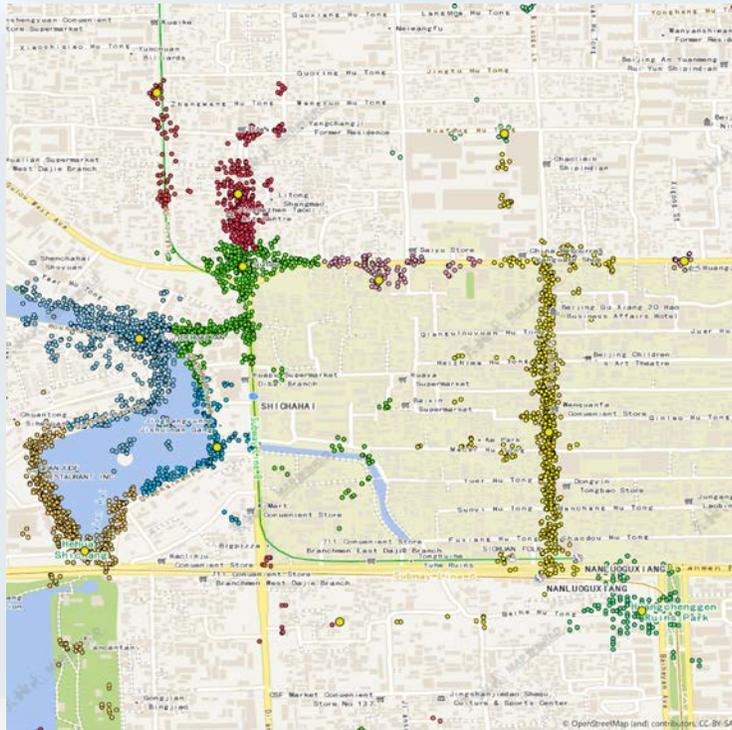
每天会搜集大量的新闻，然后把它们全部聚类，就会自动分成几十个不同的组（比如商业，科技，体育……），每个组内新闻都具有相似的内容结构



# 聚类案例

## 景区提取

对大量游客的微博定位点进行聚类，能够自动提取出不同的景点的位置分布



# 聚类案例

## 鸡尾酒会问题

在一个满是人的房间中，人们都在互相对话。我们记录下房间中的声音，利用非监督学习算法就能够识别房间中某一个人所说的话。

### Cocktail party problem

