



## 09 回归问题

西安科技大学 牟琦  
muqi@xust.edu.cn



## 9.1 机器学习基础

### 看云识天气

晴天云



非晴天云层变化



日晕三更雨，月晕午时风  
有雨山戴帽，无雨云拦腰  
朝霞不出门，晚霞行千里

早晨东云长，有雨不过晌  
早晨云挡坝，三天有雨下  
早晨浮云走，午后晒死人  
空中鱼鳞天，不雨也风颠  
天上豆荚云，不久雨将临  
天上铁砧云，很快大雨淋

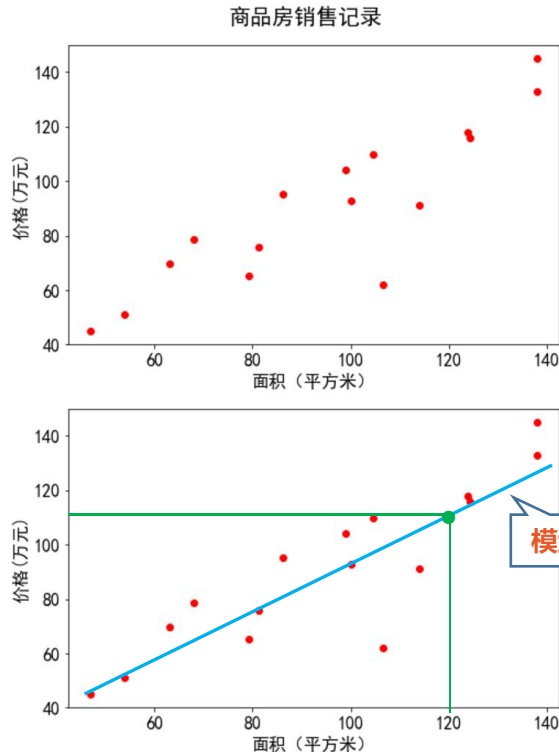


### 机器学习：从数据中学习

- 建立模型
- 学习模型
- 预测房价

#### 例：房屋销售记录

序号	面积 (平方米)	销售价格 (万元)	序号	面积 (平方米)	销售价格 (万元)
1	137.97	145.00	9	106.69	62.00
2	104.50	110.00	10	138.05	133.00
3	100.00	93.00	11	53.75	51.00
4	124.32	116.00	12	46.91	45.00
5	79.20	65.32	13	68.00	78.50
6	99.00	104.00	14	63.02	69.65
7	124.00	118.00	15	81.26	75.69
8	114.00	91.00	16	86.21	95.30



学习算法：从数据中产生模型的算法



### □ 监督学习 (Supervised Learning)

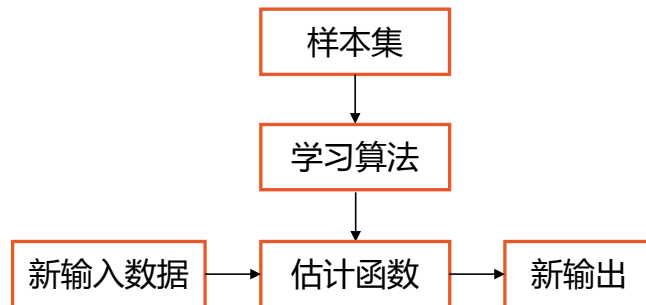
通过对**数据**的学习,寻找**属性**和**标记**之间的映射关系

#### ■ 数据集 (data set) / 样本集 (sample set)

- 示例 (instance) / 样本 (sample)
- 属性 (attribute) / 特征 (feature)
- 标记/标签 (label)

#### ■ 模型/假设/估计函数/学习器 (learner)

#### ■ 真相/真实 (ground truth)



**回归** (regression) : 预测**连续**值

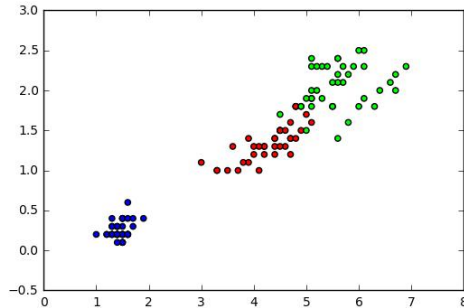
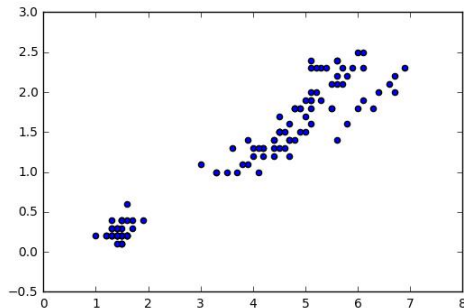
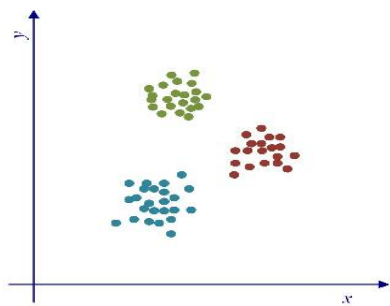
**分类** (classification) : 预测**离散**值



### □ 无监督学习 (Unsupervised Learning)

在样本数据**没有标记**的情况下，挖掘出**数据**内部蕴含的**关系**

聚类：把**相似度**高的样本聚合在一起



### □ 半监督学习 (Semi-Supervised Learning)

将有监督学习和无监督学习相结合，综合使用**大量的没有标记数据**和**少量有标记的数据**共同进行学习



### 机器学习的发展和应用

