

实例: 输入的每类值.

样例: 每种实例的集合.

训练集: 用于进行训练, 产生模型或算法的数据集.

测试集: 用来测试所得模型或算法的数据集.

特征向量: 属性的集合, 用向量来表示一个实例, 其附属于一个实例.

标记: 对实例的类别进行区别标记, 即最后所输出的结果的标记.

正例 \rightarrow 是, 1 **反例** \rightarrow 否, 0

分类问题: 目标标记为类别型的数据.

回归问题: 目标标记为连续型的数据 (如房价预测)

有监督学习: 训练集有类别标记.

无监督学习: 训练集无类别标记.

半监督学习: 训练集中部分有类别标记, 部分没有.

机器学习步骤

① 将数据集分为训练集、测试集.

② 用训练集及其特征向量来训练出算法或模型.

③ 用测试集来评估算法, 并做出调整.