

## Ch-01 绪论

令  $D = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m\}$  表示包含  $m$  个示例的数据集，每个示例由  $d$  个属性描述，则每个示例  $\vec{x}_i = \{x_{i1}; x_{i2}; \dots; x_{id}\}$  是  $d$  维样本空间  $\mathcal{X}$  中的一个向量， $\vec{x}_i \in \mathcal{X}$ ，其中  $x_{ij}$  是  $\vec{x}_i$  在第  $j$  个属性上的取值， $d$  称为样本  $\vec{x}_i$  的维数。

当预测任务是离散值，则称为分类(classification)，若预测任务是连续值，则称为回归(regression)。根据训练数据是否拥有标记信息，学习任务可分为监督学习(supervised learning)和无监督学习(unsupervised learning)两类，分类和回归是前者的代表，聚类是后者的代表。

学得模型适用于新样本的能力，被称为泛化(generalization)能力，具有强泛化能力的模型能够很好地适用于整个样本空间。我们假设样本空间中全体样本服从一个分布  $\mathcal{D}$ ，我们获得的每个样本都是独立地从这个分布上采样获得的，即独立同分布(independent and identically distributed)。

归纳和推理是科学推理的两大基本手段，归纳学习有狭义和广义之分，广义的归纳学习相当于从样例中学习，而狭义的归纳学习则要求从训练数据中习得概念，因此也称为概念学习。

机器学习算法在学习过程中对某种类型假设的偏好，被称为归纳偏好。奥卡姆剃刀(Occam's razor)原则为若有多个假设与观察一致，选择最简单的那个。

没有免费的午餐 (No Free Lunch) 定理：所有学习算法的期望性能相同。