

监督学习

监督学习(supervised learning)的任务是学习一个模型，使模型能够对任意给定的输入，对其相应的输出做一个好的预测。用户将成对的输入和预期输出数据提供给算法，算法从中找到一种方法（具体方法不用深究），然后根据给定输入给出预期输出。

1. 空间的概念

输入空间：每个具体的输入是一个实例，通常由特征向量(feature vector)表示，所有特征向量存在的空间被称为特征空间(feature space)，特征空间的每一维对应一个特征

输出空间

2. 联合概率分布

监督学习假设输入与输出的随机变量 X 和 Y 遵循联合概率分布 $P(X,Y)$ 。 $P(X,Y)$ 表示分布函数，或分布密度函数。在学习过程中假设这一联合概率分布存在，但是对于学习系统来说，联合概率分布的具体定义是未知的。训练数据与测试数据被看作是依联合概率分布 $P(X,Y)$ 独立同分布产生的。统计学习假设数据存在一定的统计规律， X 和 Y 具有联合概率分布的假设就是监督学习关于数据的基本假设。

监督学习的目的在于学习一个由输入到输出的映射，也可以这么说，学习的目的就在于找到更好的模型。监督学习的模型可以是概率模型也可以是非概率模型，由条件概率分布 $P(Y|X)$ 或者决策函数 $Y=f(X)$ 表示

监督学习的两个大分类：回归(Regression)和分类(Classification)。

回归问题

回归用于预测输入变量(自变量)和输出变量(因变量)之间的关系，特别是当输入变量的值发生变化时，输出变量的值也随之发生变化。回归模型正是表示从输入变量到输出变量之间映射的函数。回归问题的学习等价于函数拟合:选择一条函数曲线使其很好地拟合已知数据且很好地预测未知数据。

回归学习最常用的损失函数是平方损失函数，在此之下，回归问题的求解方法用的最多的是最小二乘法。

分类与回归的区别

- 分类是预测离散类标签的任务。
- 回归是预测连续数量的任务。
- 使用准确度评估分类预测，而回归预测则不可以
- 使用均方误差来评估回归预测，而分类预测则不能
- 回归分析用在神经网络上，其最上层是不需要加上softmax函数的，而是直接对前一层累加即可。回归是对真实值的一种逼近预测

分类算法可以预测连续值，但是连续值是类标签的概率的形式。

回归算法可以预测离散值，但是以整数量的形式预测离散值。

分类问题

分类问题，就比如我们用机器学习算法，将病人的检查结果分为有病和健康，是一个医学方面的二分类问题（将要区分的数据分为两个类别）。再例如在电子邮箱中，收到邮件之后，电子邮箱会将我们的邮件分为广告邮件，垃圾邮件和正常邮件，这就是一个多分类的问题（将要区分的数据分为多个类别）

逻辑回归算法

用回归的思路解决分类的问题

参考: [监督学习\(supervised learning\)的介绍 - 知乎 \(zhihu.com\)](#)

[回归问题 - 知乎 \(zhihu.com\)](#)

