

概述

- 机器学习
 - 监督学习
 - 无监督学习
 - 强化学习

监督学习

学习把输入映射到输出的算法，监督学习的“监督”在于为其提供**包括输入和输出**的训练数据，即给定输入 x 的正确标签 y 。

e.g.: 语音转文字，机器翻译，垃圾信息分类，广告推送

- 回归 regression
- 分类 classification

线性回归 linear regression

为一组数据拟合出一条直线，使得直线尽可能接近所有的数据点。

线性回归的“线性”指的是：模型的输出是**输入的线性组合**，而非简单的直线，即：

$$f(x) = f(x_1, x_2, \dots, x_n) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

成本函数/损失函数 cost/loss function

衡量模型预测值与真实值之间的差距。

预测值 $\hat{y}^{(i)} = f_{w,b}(x^{(i)})$ 误差 $\hat{y}^{(i)} - y^{(i)}$

平方误差项 $(\hat{y}^{(i)} - y^{(i)})^2$

总方误差 $\sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2$

均方误差 $J(w, b) = \frac{1}{m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2$

通常令均方误差为 $\frac{1}{2m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2$ ，以便于后续的求导运算

梯度下降 gradient descent

梯度下降最终会找到损失函数的局部极小值，即其局部最优解

需要不断使用不同的初始参数，最终找到全局最优解

学习率 $\alpha \in (0, 1)$ ，用于控制更新参数的幅度

梯度下降：

- $w = w - \alpha \frac{\partial}{\partial w} J(w, b)$
- $b = b - \alpha \frac{\partial}{\partial b} J(w, b)$
- 参数的梯度下降计算应**同时进行**，而不是先后进行，即先基于当前的参数计算更新的值，全部计算完成后再进行参数的更新

无监督学习

- 聚类 clustering
 - 将没有标签的数据试图分类到不同的群组中。
- 异常检测 anomaly detection
 - 检测异常事件，如诈骗
- 降维 dimensionality reduction
 - 将大数据集在损失尽可能少信息的情况下压缩为小数据集