

机器学习

- 概念：机器学习是从示例中学习的软件。你不需要编写机器学习的算法，而是通过提供大量的相关数据，来训练它们。
- 不同风格的算法：
 - 监督学习：就是你使用带有标签数据的训练算法，将一组输入对象(通常为矢量)映射到一组期望的输出值中(也称为监督信号)。你提供给监督学习的算法样本越多，其分析新数据的能力就越精确。而这也是监督学习的主要挑战。创建带有标记样本的大数据非常耗时，同时也需要大量的人力。
 - 无监督学习：“无监督学习”中，机器学习的另一个分支，则没有参考数据，一切都没有标签。换句话说，你提供输入，但不提供输出。该算法对未标记的数据进行整理，提取推论并找出模式。无人监督学习对于人类无法定义的，隐藏模式的情况尤其有用。
 - 强化学习：在强化学习中，程序员定义了状态，期望目标，允许的行动以及约束。该算法通过尝试将各种不同的允许行动进行组合来了解如何实现目标。当你知道目标是什么，但无法定义达到目标的路径时，此方法特别有效。

深度学习

- 概念：机器学习是人工智能的一个分支，但深度学习是机器学习的专门分支。深度学习使用神经网络，一个具有人类大脑结构和功能的复制品。
- 原理：深度学习算法在获得更多数据的同时，其性能也在不断改进。深度学习算法并不是直接通过将输入映射到输出的方式，而是依赖于几层处理单元。每个层将其输出传递到下一个层，进行处理，然后再传递到下一层。

监督学习（有输入，有输出）

- 通过监督给定的正确数据，从而对输入的数据进行预测的输出
- 监督分为两类：
 - 回归问题：给定一组连续的数据，从而预判新输入的数据对应的输出
 - 分类问题：给定一组离散的数据，在给定的1~∞多个特征变量中，预判输入的数据对应的输出

无监督学习（有输入，无输出）

- 给定的数据无任何特征，让机器来进行识别与学习，从而自动对数据进行处理。
- 典型算法：
 - 聚类算法：
 - 将几堆纯数据进行分类，打上标签
 - 鸡尾酒算法
 - 将两段具有重合音频的录音分别进行过滤处理，从而使每一段录音只剩下其原本该拥有的录音。

单变量线性回归

代价函数1

- m: 训练集的数量
- x: 输入变量/特征变量
- y: 输出变量/目标变量
- $(x^{(i)}, y^{(i)})$: 表示某一对数据
- θ_i : 模型参数
- $h_{\theta}(x) = \theta_0 + \theta_1 x$: 假设函数
- 简化后的代价函数:
 - ◻
- 简化后的优化目标
 - ◻
- Tips:
 - $1/2m$ 可以使数学结果更加直白,但不会影响 θ_0 和 θ_1 的计算
 - 需要求得就是使代价函数最小的 θ_0 和 θ_1 的值
- 解使得优化目标最小化的 θ_1 的过程
 - 做出 $J(\theta_1)$ 与 θ_1 的函数图
 - 通过取不同的 θ_1 ,从而求得不同的 $J(\theta_1)$ 的结果,最后使 $J(\theta_1)$ 最小的值便为所求的值
 - 这个过程实际上就是在寻求最适合的拟合假设函数 $h(\theta_1)$,从而使我们的预测结果更加精准

代价函数2

- 等高线图(考虑 θ_0 和 θ_1 ,变成三维函数图)
 - ◻

梯度下降算法

- 作用:对于任意一个函数(≥ 1 个自变量的函数),可以寻求其全局最优解或者局部最优解
- α : 学习效率,为一个正数
- $dJ(\theta_0, \theta_1)/d\theta_j$ 表示J对 θ_j 的导数,即代价函数在 θ_j 这个点的斜率

- 该算法不断迭代,更新 θ_j 的值,直到更新到 $J(\theta_j)$ 极小值的点
 -
- 模拟图像
 - □
- Tips
 - α 不宜过大.否则可能变化梯度太大,从而无法到达极小值点
 - α 不宜过小,否则需要迭代更新 θ_j 太多次,才能到达极小值点
 - 当导数等于0时, θ_j 将不再更新.

线性回归的梯度下降

- 背景介绍:
 - 通过Batch梯度下降算法寻求线性回归代价函数的极小值
 - 工具:
 - □
 - α 后半部分通过微积分的偏导推算而来的
 - 由于线性回归的代价函数使凸函数,所以一定拥有最优解
 - □

线性代数回归

矩阵和向量

- 矩阵:
 - □
- 向量（特殊矩阵， $n \times 1$ ）:
 - □

矩阵向量乘法

- □
- 线性回归预测求解技巧（一个假设函数）:
 - □
 - $\text{prediction} = \text{DataMatrix} * \text{parameters}$

矩阵乘法

- □
- 线性回归求解技巧（多个假设函数）:
 - □

矩阵乘法的特征

- 单位矩阵:
 - □
 - Tip:
 - 任何矩阵x单位矩阵，都保持不变
 - 确保维度相同即 $m \times n$ 单位矩阵为 $m \times n$ （右乘单位矩阵）或者 $m \times m$ （左乘单位矩阵）
- 结合律:
 - $(AM)N = A(MN)$
- notes:
 - $AB \neq BA$
 - $|A| = |A|$ (I 为单位矩阵)

逆和转置

- $A * A^{-1} = I$
 - A 的维度必须为 $n \times n$
- 转置:每个元素的行和列交换
 - $A_{ij} = B_{ji}$
 - notes:可作四十五度对角线,做镜像交换

多变量线性回归

- 多元梯度下降法
 -
- 特征缩放
 -

- notes:
 - u_1 为 x_1 样本特征的平均值
 - s_1 为 $\text{Max}(x_1)-\text{Min}(x_1)$
 - 若 $-1 \leq x_1 \leq 1$,或者在这周围,那么不用进行缩放
- 学习率选择:
 - 目标函数随迭代次数增加的曲线图

□

- 几种学习率选择正确导致的曲线图

□

- notes:
- 通常出现上述情况,都是选择降低 α 的值,已知 α 足够小时, $J(\theta)$ 一定可以取得最优解

- □
 - Notes:以3倍为基础逐渐升高 α ,一般取比最大值略小的 α