

# 吴恩达机器学习

---

## 1 绪论：初识机器学习

### 1.1 应用领域

- 数据挖掘
  - 网页点击数据；
  - 医疗记录；
  - 生物工程等数据集；
- 无法编写的程序
  - 自动驾驶；
  - 手写体识别；
  - 自然语言处理；
  - 计算机视觉；
- 私人定制程序
  - 智能推荐；
- 理解人类的学习过程和大脑

### 1.2 什么是机器学习

- Arthur Samuel在1959年的定义：在没有明确设置的情况下，是计算机具有学习能力的领域；
- Tom Mitchell在1998年的定义：计算机程序从经验E中学习，解决某一项任务T，进行某一性能度量P，通过P测定在T上的表现因经验而提高；

### • 1.3 监督学习和无监督学习

- 监督学习 (Supervised Learning)
  - 根据给定的算法和数据集，且该数据集包含了正确的答案，便可以对应给出任意给定一个数据的对应答案，即解决的是回归问题或者分类问题；
- 无监督学习 (Unsupervised Learning)
  - 给算法大量的数据，要求其找出数据结构，即解决的是聚类问题；

## 2 单变量线性回归

### 2.1 代价函数

- 假设函数

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

- 参数

$$\theta_0, \theta_1$$

- 代价函数

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

- 目标

$$\underset{\theta_0, \theta_1}{\text{minimize}} J(\theta_0, \theta_1)$$

说明：此处之所以是  $\frac{1}{2m}$ ，原因是在按梯度下降法求最小值点时刚好可以把平方的2约掉，从而变为  $\frac{1}{m}$ ，且不论是  $\frac{1}{m}$ ，还是  $\frac{1}{2m}$ ，最后计算得出的  $\theta_0$ 、 $\theta_1$  都是一样的。

## 2.2 梯度下降法

- 算法描述

$$\begin{aligned} & \text{repeat until convergence} \{ \\ & \quad \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0 \text{ and } j = 1) \\ & \} \end{aligned}$$

说明：此处的“:=”表示赋值运算。

- 同步更新

$$\begin{aligned} \text{temp0} &:= \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\ \text{temp1} &:= \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \\ \theta_0 &:= \text{temp0} \\ \theta_1 &:= \text{temp1} \end{aligned}$$

- $J(\theta_0, \theta_1)$  的偏导数

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

## 3 线性代数回顾

### 3.1 需要用到的线性代数基础知识