

机器学习笔记一——梯度下降和最小二乘法

本笔记是在斯坦福大学的机器学习公开课上学习总结的，纯属个人观点。

首先来一个例子，假设我们需要对本市的房价进行建模，我们已知的变量是房间的占地面积，那么自变量是房间的占地面积，因变量是房子的价格，例如我们可能有这样的数据：

房子面积 m*m	价格 RMB(w)
100	300
120	380
235	500
48	120

以上数据是自己编造的，仅为说明问题，并无科学依据。

那么我们希望得到已知房子面积，对房子价格的一个可靠的估计，更一般的讲，我们可能收集到影响房间价格的其他因素，例如房间个数、地理位置等，我们希望利用这些数据对房间价格进行估计，这类问题成为回归（regression problem），首先需要统一一下符号：

$X_i = \{x_0, x_1, x_2, \dots, x_n\}$ 输入变量，在这个例子中可能是房子面积、房间个数等。

n 输入变量的维数。

Y_i 输出变量，在这个例子中为房子的价格。

$h(\theta) = \sum_{i=0}^n \theta_i x_i$ 我们的模型，在这里我们假设 $x_0 = 1$

$\theta = \{\theta_0, \theta_1, \dots, \theta_n\}$ 参数，需要确定的东西。

线性回归的一般模型表示为： $h_x(\theta) = \sum_{i=0}^n \theta_i x_i = \theta^T X$

我们希望上述模型是对 Y_i 的一个很好的估计，因此定义如下的错误估计函数：

$$J_\theta = \frac{1}{2} \sum_{i=0}^m (h_x^i(\theta) - Y_i)^2$$

上式中，为了对数据集进行求和，对 $h_x(\theta)$ 加了上标，表示一个样本的情况。此问题为典型的最小二乘问题，在这里首先用梯度法进行求解，然后利用最小二乘解法求解。

一、梯度下降法

梯度下降法又名最快下山算法，其基本思想是对于连续“曲面”上的点，计算该点的梯度，梯度方向即为下降最快的方向，此算法为局部最优算法，其模型为：

$$\theta := \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

对于每一个 θ_i 为:

$$\begin{aligned} & \frac{\partial J(\theta)}{\partial \theta_i} \\ &= \frac{\partial}{\partial \theta_i} \frac{1}{2} (h_\theta(x) - y)^2 \\ &= 2 \cdot \frac{1}{2} (h_\theta(x) - y) \frac{\partial}{\partial \theta_i} (h_\theta(x) - y) \\ &= (h_\theta(x) - y) x_i \end{aligned}$$

因此

$$\theta_i := \theta_i - \alpha (h_\theta(x) - y) x_i$$

这就是梯度下降法，值得注意的是梯度下降法对于回归问题非常有用，可以证明线性回归问题是一个凸优化问题，只有一个局部最优解即全局解。

α 决定了下降的速度。

二、最小二乘的数学求解

假设 feature 向量组成以下的矩阵:

$$A = \begin{pmatrix} \cdots & x_{1,i} & \cdots \\ \cdots & x_{2,i} & \cdots \\ \cdots & x_{m,i} & \cdots \end{pmatrix}$$

那么，参数向量表示为:

$$X = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_n \end{pmatrix}$$

那么

$h_x(\theta) = AX$ 作为 Y 的一个估计，我们的目标是求解 X 。一般 A 的行数要远远大于列数，即要有很多的样本数据。

$$p = AX - Y = \begin{pmatrix} h_x(\theta_1) - y_1 \\ h_x(\theta_2) - y_2 \\ h_x(\theta_m) - y_m \end{pmatrix}$$

则:

$$J = p^T p = (AX - Y)^T (AX - Y)$$

通过对 J 求梯度向量可以得到

$$0 = \nabla J = 2A^T AX - 2A^T Y$$

所以: $A^T A X = A^T Y$

假设 A 为 $m \times n$ 的矩阵, 则 $A^T A$ 的维数为 $n \times n$, 可以证明在 A 行满秩的条件下假设 $B = A^T A$, 则 B 为正定矩阵

$BX = A^T Y \rightarrow X = B^{-1} A^T Y$ 即为该方程的最小二乘解, 具体求解方法, 学过的有直接逆矩阵、LU 分解、QR 分解等。