线性回归

- 线性回归是回归问题，通过给定训练集m，选择一个模型函数h，通过计算得到模型函数的最优解，并计算最优解下的参数。
- 线性回归分为：
  - 单变量线性回归：

$$h(x) = \theta_0 + \theta_1 x_1$$

  - 多变量线性回归：

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

- 通用表达式：

$$h(x) = \theta^T X = \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

- 代价函数：代价函数是计算建立模型和真实数据的误差。

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

  - m：训练样本个数
  - n：训练集的特征个数

正规方程

- 正规方程是通过求解下面的方程来找出使得代价函数最小的参数的：

$$\frac{\partial}{\partial \theta_j} J(\theta_j) = 0$$

  用来求解θ的最小值。

$$\theta = (X^T X)^{-1} X^T y$$

  - x：训练集矩阵：(mx(n+1))
  - y：训练集对应的正确答案：(1xm)
- 注意：矩阵不可逆不能使用正规方程（原因：1.各个特征不独立 2.特征数量大于样本个数）

梯度下降算法（局部最小值）

- 梯度下降算法是计算达到最优解时，代价函数最小值时的θ值。`每次选取所有样本更新θ`，α指学习率。

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_0, \ldots, \theta_n)$$

- 代入J(θ)，得

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X^{(i)}) - y^{(i)})^2$$

- 求导后得到

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(X^{(i)}) - y^{(i)})$$

- 每个θj必须是同步变换。

特征缩放

- 对于多个特征值相差较大时，需要特征缩放，把所有参数缩放到-1~1的范围，让x适应每个参数。也就是

$$X_n = \frac{X_n - \mu_n}{S_n}$$

  - μ是平均值，S是标准差。

随机梯度下降算法

- 每次只取一个样本更新θ。
- 常用学习率α：`0.01`、`0.03`、`0.1`、`0.3`、`1`、`3`、`10`。
- 梯度下降算法与正规方程的比较：

梯度下降算法	正规方程
需要选择学习率α	不需要
需多次迭代	一次运算得出
特征数量n较大时也适用	n ≤ 10000
适用于各种模型	只适用于线性模型

- 对学习率选取时的要求：如果学习率过大，则每次执行梯度下降可能跳过最小值，呈现震荡曲线；如果学习率选择过小，则每次迭代比较慢，最后可能会出现过拟合的现象。
- 如果θ初始值在最小值的位置，则损失函数是0，不会变化。