回归

2023年4月6日　　下午 10:41

回歸問題是指預測一個連續值的問題，例如房價預測、股票價格預測等。在回歸問題中，我們通常會使用一些特徵來預測目標變量，這些特徵可以是數值型、類別型或是其他形式的。我們可以使用各種回歸模型來建立特徵和目標變量之間的關係，例如線性回歸、多項式回歸、決策樹回歸、隨機森林回歸等。

回归分析指研究一组随机变量(Y1 ，Y2 ，…，Yi)和另一组(X1，X2，…，Xk)变量之间关系的统计分析方法，又称多重回归分析。

一、线性回归（预测连续值数量）

最小二乘法：

它通过最小化误差的平方和寻找数据的最佳函数匹配。

线性回归的目标是最小化代价函数

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

假设函数$h\theta(x)$由线性模型$h_\theta(x) = \theta^T x = \theta_0 + \theta_1 x$给出。

模型的参数θ_j是需要被调整，从而使代价$J(\theta)$最小化的值。一种方法是使用批处理梯度下降算法(batch gradient descent algorithm)。在批次梯度下降中，每次迭代都会执行更新

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \quad (\text{simultaneously update } \theta_j \text{ for all } j)$$

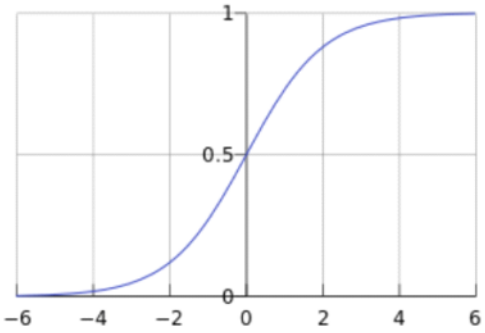
随着梯度下降的每一步，参数θ_j将接近最佳值，从而实现代价$J(\theta)$的最低。

二、逻辑回归（预测离散变量，二分类算法）

Sigmoid函数，也称为logistic函数：

- $g(z) = \frac{1}{1+e^{-z}}$

其函数曲线如下：



从上图可以看到sigmoid函数是一个s形的曲线，它的取值在[0, 1]之间，在远离0的地方函数的值会很快接近0或者1。它的这个特性对于解决二分类问题十分重要

逻辑回归的假设函数形式如下：

- $h_\theta(x) = g(\theta^T x), g(z) = \frac{1}{1+e^{-z}}$

所以：

- $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

其中 x 是我们的输入， θ 为我们要求取的参数。

一个机器学习的模型，实际上是把决策函数限定在某一组条件下，这组限定条件就决定了模型的假设空间。当然，我们还希望这组限定条件简单而合理。而逻辑回归模型所做的假设是：

- $P(y = 1|x; \theta) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$

这个函数的意思就是在给定 x 和 θ 的条件下 $y = 1$ 的概率。

三、Softmax回归（多分类）

Softmax 回归是 Logistic 回归的一般形式，logistic 回归是 softmax 回归在 k=2 时的特殊形式。逻辑回归通常用作2类的分类器，softmax则用作多类的分类器。

对于输入数据$\{(x_1,y_1),(x_2,y_2),\ldots,(x_m,y_m)\}$有 k 个类别，即$y_i \in \{1,2,\ldots,k\}$，那么 softmax 回归主要估算输入数据 x_i 归属于每一类的概率，即

$$h_\theta(x_i) = \begin{bmatrix} p(y_i=1|x_i;\theta) \\ p(y_i=2|x_i;\theta) \\ \vdots \\ p(y_i=k|x_i;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ \vdots \\ e^{\theta_k^T x_i} \end{bmatrix} \tag{1}$$

其中，$\theta_1,\theta_2,\ldots,\theta_k \in \theta$是模型的参数，乘以$\frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_i}}$是为了让概率位于[0,1]并且概率之和为1，softmax 回归将输入数据 x_i 归属于类别 j 的概率为

$$p(y_i=j|x_i;\theta) = \frac{e^{\theta_j^T x_i}}{\sum_{l=1}^{k} e^{\theta_l^T x_i}} \tag{2}$$