

线性回归

多元线性回归

基本概念

线性回归是机器学习中有监督机器学习下的一种算法。回归问题主要关注的是因变量（需要预测的值，可以是一个也可以是多个）和一个或多个数值型的自变量（预测变量）之间的关系。

需要预测的值：即目标变量，target， y ，连续值的预测

隐形目标变量的因素： $X_1 \dots X_n$ ，可以是连续值也可以是离散值

因变量和自变量之间的关系：即模型，model，也就是我们要求解的

连续值

在某个范围内可以取任意值的数值

身高、体重...

离散值

孤立的点集，像区间，它在每一点上都是连续的，而像整数集，它的每一元素之间都有一点的距离。

简单线性回归

前面提到过，算法说白了就是公式，简单的线性回归属于一个算法，它所对应的公式

$$y = wx + b$$

这个公式中， y 是目标变量即未来要预测的值， x 是影响 y 的因素， w 、 b 是公式上的参数即要求的模型，其实 b 就是截距， w 是斜率！所以很明显如果模型求出来，未来影响 y 值的未知数就是一个 x 值，也可以说影响 y 值的因素只有一个，所以这就是叫简单线性回归的原因。

同时可以发现从 x 到 y 的计算， x 只是一次方，所以这叫做线性回归的原因。人工智能要求算法是最优解！

最优解

Actual value：真实值，一般用 y 表示

Predicted value：预测值，是把已知 x 带入公式里面和猜出来的参数 w 、 b 计算得到，一般使用 \hat{y} 表示

Error：误差，预测值与真实值的差距，一般使用 ϵ 表示

最优解：尽可能地找到一个模型使得整体的误差最小，整体的误差通常叫做 Loss

Loss：整体的误差，Loss通过损失函数 Loss function(最小二乘法) 计算得到

多元线性回归（复杂）

现实生活中，往往影响结果的 y 的因素不止一个，这时 x 就从一个变成了 n 个， $X_1 \dots X_n$ 同时简单线性回归的公式也就不再适用了。多元线性回归公式如下：

$$\hat{y} = w_1 X_1 + w_2 X_2 + \dots + w_n X_n + b \quad (1)$$

b 是截距，也是使用 w_0 来表示 (2)

$$\hat{y} = w_1 X_1 + w_2 X_2 + \dots + w_n X_n + w_0 \quad (3)$$

$$\hat{y} = w_1 X_1 + w_2 X_2 + \dots + w_n X_n + w_0 * 1 \quad (4)$$

使用向量来表示， x 表示所有的变量，是一维向量 w 表示所有的系数（包含 w_0 ），是唯一向量，根据向量乘法规律，可以这么写：

$$\hat{y} = W^T X \quad (5)$$

T 就是转置， W 是行，那么转置后就是列