第 3 章 线性模型

一、问题提出

设某类数据样本 x，每个样本有 d 维属性。若有 N 个这样的样本，则可以表示为矩阵形式：

$$\mathcal{X} = \begin{bmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nd} \end{bmatrix}$$

另外，对应上面的 N 个样本，我们给出了 N 个标签，记为 ：

$$\mathcal{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

参考书上表 1.1(第 4 页)的西瓜数据集，可以写出对应的$\mathcal{X}$与$\mathcal{Y}$。

我们的目标，是基于上面的数据与标签，学习到一个函数$\mathcal{F}$，使其最好地拟合已知的$\mathcal{X}$ -> $\mathcal{Y}$ 所描述的隐含联系。为我们后面，在拿到一个新的样本以后，可以较好地知道该为它贴什么标签。

二、线性模型

线性关系，严格地说，是指线性变换，它满足“在两个向量空间之间的函数，它保持向量加法和标量乘法的运算”（百度百科）

即

$\mathcal{F}$(X + Y) = $\mathcal{F}$(X) + $\mathcal{F}$(Y)　　X, Y 为向量　　(加性)

$\mathcal{F}$(aX) = a$\mathcal{F}$(X)　　a 为标量　　(乘性)

例如，y = ax，就是x -> y的一个线性变换

验证如下：

令y = f(x) = ax

则

f($x_1 + x_2$) = a($x_1 + x_2$) = ax_1+ ax_2 = f(x_1) + f(x_1)

f(kx) = a(kx) = kax = kf(x)

通常y = ax + b 不是严格意义上的线性变换，对于偏置项 b 的影响，在图形上需要

用仿射变换来完成。这里不就再展开了。在笛卡尔坐标系中，二维平面上y = ax + b 表示的是一条直线；三维空间中，y = ax + b 则是一个平面；更高维的空间，被称为超平面。

线性模型，就是用y = ax + b 的函数关系，去描述数据与标签之间的映射关系。对于

1. 回归问题，就是用y = ax + b 去预测，在提供了新样本x后，应该用什么标签y去对应。图形上看，就是用一条直线或平面（超平面）去近似x，y之间的关联。

2. 分类问题，就是用y = ax + b 去把新样本x，分到不同的类别里去，这个类别的定义，就是y的标签来决定了。图形上看，就是用一条直线或平面（超平面）去把不同的样本点分开。

三、学习算法：线性回归

模型选好了，就需要确定对应的参数，即y = ax + b 中的 a 和 b 应该是什么。这就需要用算法，把数据作为输入，计算出 a 和 b 的值。表示成矩阵的表示，就是 f(x) = $\mathcal{W}^T x$ + b 即(书上 3.2 式，第 53 页)

1. 一元线性回归

先讨论的是一元的情况，即数据的属性只有一个，取值为实数。这就是我们熟悉的 f(x) = wx + b 的情况。这里样本的个数为 m。为了求得 w 与 b 的值，或者是最理想值，一个方法是取均方误差(Mean Squared Error)最小。即对于每个样本，从 1 到 m，我们计算各种 w 与 b 的可能值，直到下式取值最小：

MSE = $\sum_{i=1}^{m}(wx_i + b - y_i)^2$　　（书上式 3.4, 第 54 页）

这个公式的思想是，以x_i为一个样本，通过计算 wx_i+b 得到一个估计的y_i^*，将它与原始x_i的标签y_i去做差，得到的是一个误差，平方后即得到一个正数，可以理解为估计值与真实值之间距离的一个度量。然后把当前 w 和 b 取值下的全部样本的误差求和，即在假设空间中遍历 w 与 b 的可能性，找到 MSE 最小，这个 w, b 组合，就是我们想要的参数。

遍历假设空间是基本思想，但求 MSE 的最小，有直接的方法。因为 MSE 被构造成了一个 w 和 b 的二次函数。我们知道，一元二次函数是一个抛物线，开口向上，有最小值。二元二次函数是一个碗状的曲面，这里也正好是碗口向上，有最小值。且这个函数好就好在容易求导，其导数等于 0 的位置就是最小值点。故直接对 MSE 对 w 和 b 分别求导，会得到两个方程，两个未知量，可以求解了。

$$\frac{\partial E}{\partial w} = \sum_{i=1}^{m} 2(wx_i + b - y_i)x_i = 0$$

$$2\sum_{i=1}^{m}(wx_i^2 + (b - y_i)x_i) = 0$$

$$w\sum_{i=1}^{m}x_i^2 + \sum_{i=1}^{m}x_i(b - y_i) = 0$$

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}x_i(y_i - b)$$

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}x_iy_i - \sum_{i=1}^{m}x_ib \qquad ①$$

同理，对 b 求导：

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{m}2(wx_i + b - y_i) = 0$$

$$\sum_{i=1}^{m}(wx_i - y_i) = -\sum_{i=1}^{m}b$$

$$mb = \sum_{i=1}^{m}(y_i - wx_i)$$

$$b = \frac{1}{m}\sum_{i=1}^{m}(y_i - wx_i) \qquad ②$$

将②式代入①式，可以解得：

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}x_iy_i - \sum_{i=1}^{m}x_i(\frac{1}{m}\sum_{i=1}^{m}(y_i - wx_i))$$

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}x_iy_i - \sum_{i=1}^{m}x_i(\frac{1}{m}\sum_{i=1}^{m}y_i - \frac{1}{m}\sum_{i=1}^{m}wx_i)$$

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}x_iy_i - \frac{1}{m}\sum_{i=1}^{m}x_i\sum_{i=1}^{m}y_i + \frac{w}{m}(\sum_{i=1}^{m}x_i)^2$$

由于$\frac{1}{m}\sum_{i=1}^{m}x_i = \overline{x}$(样本均值)，在给定样本数据后，它是个常数。代入上式得：

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}x_iy_i - \overline{x}\sum_{i=1}^{m}y_i + \frac{w}{m}(\sum_{i=1}^{m}x_i)^2$$

$$w = \frac{\sum_{i=1}^{m}x_iy_i - \overline{x}\sum_{i=1}^{m}y_i}{\sum_{i=1}^{m}x_i^2 - \frac{1}{m}(\sum_{i=1}^{m}x_i)^2}$$

$$w = \frac{\sum_{i=1}^{m} y_i(x_i - \bar{x})}{\sum_{i=1}^{m} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2}$$

这里，得到书上式 3.7，第 54 页。

2. 多元线性回归