下面来仔细看一下各组数据。图 2-3b 的数据看上去是沿曲线分布的，所以假定这组数据是线性分布的就不合适。图 2-3c 中有离群值，应考虑对其进行去除离群值的预处理，或者应用便于处理离群值的其他方法。图 2-3d 是对数据两两之间没有相关关系但存在离群值的数据集应用线性回归而得到回归直线的例子。对原本不遵循线性分布的数据强行进行线性回归也得不到好的结果。拿到数据之后，首先应该进行可视化，再考虑是否进行线性回归。

均方误差的最小化方法

在“算法说明”部分，我们了解了均方误差可用于评估具有不同学习参数的直线。

然而，虽然通过如表 2-3 所示的数据比较了不同的学习参数产生的误差大小，但我们还不了解具体的学习参数的计算方法。

本节将介绍如何获得使均方误差最小的学习参数。

▼表 2-3 学习参数和均方误差的关系

线性回归	w_0	w_1	均方误差
(a)	0.823	0.706	2.89
(b)	4.5	−0.125	5.83

从表 2-3 可以看出，由于学习参数的变化，作为误差函数的均方误差也会发生变化。也就是说，均方误差可以使用学习参数的函数表示：

$$L(w_0, w_1) = \frac{\sum_{i=1}^{n}\left[y_i - (w_0 + w_1 x_i)\right]^2}{n}$$

这时将表 2-2 的数据代入目标变量 y_i 和特征变量 x_i，就可以仅用 w_0、w_1 表示均方误差：

$$L(w_0, w_1) = \frac{\sum_{i=1}^{4}\left[y_i - (w_0 + w_1 x_i)\right]^2}{4} = w_0^2 + 24.5w_1^2 + 9w_0w_1 - 8w_0 - 42w_0 + 21$$

该式是 w_0、w_1 的二次函数，函数图形如图 2-4 所示。从图 2-4 可以看出，当 w_0、w_1 发生变化时，误差的值也各不相同。另外，图 2-4 上标记星号的点对应于表 2-3 中的两种情况。表 2-3 中的 (a) 的学习参数与误差函数取最小值时的学习参数一致，可以看出这是数据点的最佳学习参数。