函数的极值点，可以看到，公式中 $(x^{(i)},y^{(i)})$ 和 m 都是已知的，因此可以直接代入求解。

对于单变量线性回归算法来说，一方面虽然其模型简单，仅仅包含两个变量，但是在实际中并不可能只用一个特征变量去预测目标变量，大数据时代常常要找很多特征变量来预测目标变量，即后边章节要讲的多变量线性回归；另一方面，最小二乘法是最基础的算法，很多时候要用梯度下降法等去优化，得到更优的解。因此，在这里式（3.6）的结果更多的作用是让我们理解最小二乘法的求解过程。

3.2.3　单变量线性回归的评价与预测

求解参数 θ_0、θ_1 取值之后，基于已知的特征变量 x，根据 $y = h_\theta(x) = \theta_0 + \theta_1 x$ 就可以求出目标变量 y 的预测值。利用预测值与实际值的比较，我们就可以对所使用的算法进行评价。在回归算法中，主要的评价方法有以下几种：

❑　平均绝对差值（MAE）

平均绝对差值（Mean Absolute Error，MAE）是目标变量每个样本 i 的预测值与实际值差的绝对值，加总之后求平均，其公式为：

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|yf^{(i)} - y^{(i)}| \tag{3.7}$$

其中，$yf^{(i)}$ 为预测值。

❑　均方误差（MSE）

均方误差（Mean Square Error，MSE）是目标变量每个样本 i 的预测值与实际值差的平方，加总之后求平均，其公式为：

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(yf^{(i)} - y^{(i)})^2 \tag{3.8}$$

❑　均方根误差（RMSE）

均方根误差（Root Mean Square Error，RMSE）就是在均方误差 MSE 开根号，其公式为：

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(yf^{(i)} - y^{(i)})^2} \tag{3.9}$$

不同于均方误差 MSE 的是，均方根误差的单位与所用数据的单位是相同的。另外，可以看到均方误差 MSE 与成本函数比较相似，这也是使用均

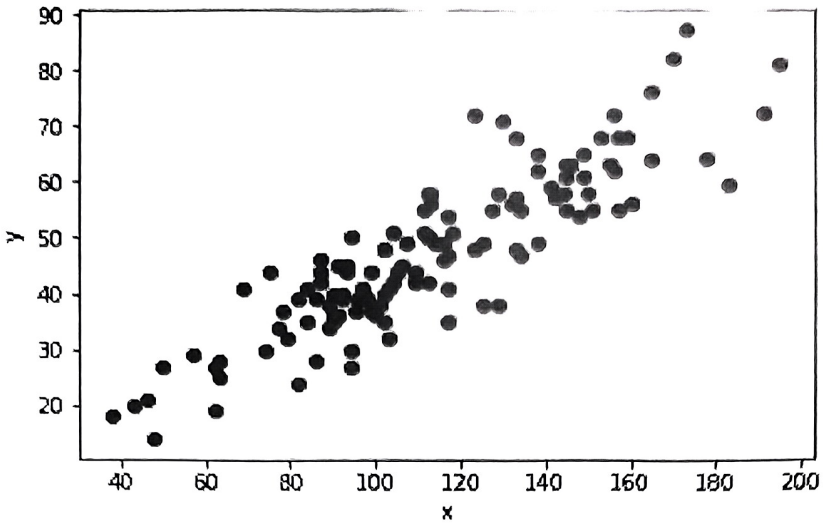

图 3.3　影厅观影人数与影厅面积的散点图

3.2.2　单变量线性回归的常规求解

对于形如：$h_\theta(x) = \theta_0 + \theta_1 x$ 的单变量线性回归算法，求解参数 θ_0、θ_1 的常规方法就是最小二乘法。从拟合的角度看，一个好的估计要求其估计结果与初始值偏差较小，也就是估计误差较小，那么我们就引入普通最小二乘估计（Ordinary Least Square Estimation，OLS）对线性回归模型进行估计。最小二乘法的基本思想是拟合线性回归直线与所有样本数据点都比较靠近，即要目标值 y_i 与其预测值的差 $y_i - h_\theta(x) = y_i - (\hat{\theta}_0 + \hat{\theta}_1 x)$ 越小越好，不同参数 θ_0、θ_1 的取值不同，目标值 y_i 与其预测值的差值是不同的，差值最小的那条线段也就是最小二乘法得到的拟合曲线，如图 3.4 所示。

因此，求解参数 θ_0，θ_1 取值的问题转化成了一个 mininize$(h_\theta(x) - y)$ 的问题，这就是最小二乘法的基本原理。在介绍最小二乘法的求解思路之前，先引入机器学习中一个常用的概念或函数，即成本函数（Cost Function，也被称为代价或损失函数），其表达式为：

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{3.3}$$

其中，m 表示训练集中实例的数量（训练集中的训练样本个数）；$(x^{(i)}, y^{(i)})$ 表示第 i 个观察实例（第 i 个训练样本，上标 i 只是一个索引，表示第 i 个训练样本，即表中的第 i 行）。例如，在影厅观影人数与影院面积的关系

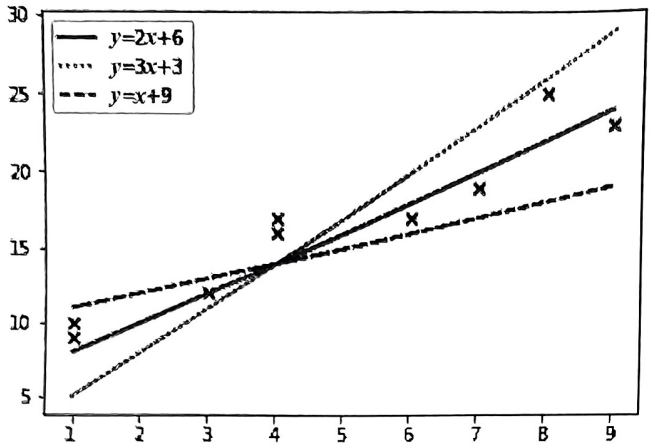

图 3.4　不同参数取值得到的拟合曲线情况

例子中，第一个训练样本的 x 值为 45、y 值为 106，第二个训练样本的 x 值为 44、y 值为 99。

成本函数（式 3.3）也被称作平方误差函数（Squared Error Function），对于回归问题来说，误差平方和函数是一个比较合适、常用的选择，当然，也可以选择一些其他形式的成本函数。最小二乘法中 mininize$(h_\theta(x) - y)$ 的问题，就可以转化为求解成本函数最小的问题。

成本函数最小问题的常规解法就是，对成本函数 $J(\theta)$ 求偏导后并令其等于零，所得到的 θ 即为模型参数的值。即：

$$\frac{\partial}{\partial\theta_k}J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_k^{(i)} = 0, \quad k = 0,1,2,\cdots,n \tag{3.4}$$

由于这里为单变量线性回归算法，因此 $k=0$，1，结合式（3.2）和式（3.4）可以转变成为一个方程组：

$$\left.\begin{aligned}\frac{\partial J(\theta)}{\partial\theta_0} &= \frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)}) \times 1 = 0\\ \frac{\partial J(\theta)}{\partial\theta_0} &= \frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})x^{(i)} = 0\end{aligned}\right\} \tag{3.5}$$

对方程组求解，可以得到：

$$\left.\begin{aligned}\theta_1 &= \frac{m\sum x^{(i)}y^{(i)} - \sum x^{(i)}\sum y^{(i)}}{m\sum x^{(i)2} - (\sum x^{(i)})^2}\\ \theta_0 &= \frac{\sum y^{(i)}}{m} - \theta_1\frac{\sum x^{(i)}}{m}\end{aligned}\right\} \tag{3.6}$$

这就是单变量线性回归算法下最小二乘法的解法，就是求得平方损失

方误差 MSE 作为评价指标的缘由。

❑ 拟合优度（R^2）

拟合优度（R^2）是判断回归模型拟合程度好坏的最常用的指标，来自于统计学中的回归评价指标。

$$R^2 = \frac{SSR}{SST} = \frac{\sum (yf^{(i)} - \overline{y^{(i)}})^2}{\sum (y^{(i)} - \overline{y^{(i)}})^2} \tag{3.10}$$

拟合优度是对回归模型拟合程度的综合度量，拟合优度越大，回归模型拟合程度越高。R^2 表示因变量 y 的总变差中可以由回归方程解释的比例。可决系数 R^2 具有非负性，取值范围为0到1，它是样本的函数，是一个统计量。其取值越接近于1，说明拟合效果越好。

除了对自身数据进行拟合与评价外，当模型训练完毕后，我们需要使用一个与训练数据集独立的新的数据集去对模型进行验证。因为模型本身就是使用训练数据集训练出来的，因此它已经对训练集进行了很好的拟合，但是它在新的数据集上的效果有待验证，因此需要使用新的与训练集独立的数据集对模型进行训练，确保该模型在新的数据集上也能够满足要求。模型对新的数据也有很好的预测的能力则被称为模型的泛化能力。

那么新的数据集如何得来呢？一般是将已有的数据集随机划分成两个部分，一部分用来训练模型，另一部分用来验证与评估模型。另一种方法是重采样，即对已有的数据集进行有放回地采样，然后将数据集随机划分成两个部分，一部分用来训练，另一部分用来验证，即训练集与测试集。关于两个数据集合如何划分的问题，在机器学习中也有很多方法，这在后边将会有所涉及。只有通过这种已有数据集的反复验证，才能确保我们的模型泛化能力较好，从而更好地用在未知目标变量数据集的预测上。

3.3 用机器学习思维构建单变量线性回归模型

本节基于上述对单变量线性回归算法原理、求解、评价等内容的介绍，从机器学习思维的角度，通过一个简单案例来构建单变量线性回归模型，并进行预测。在这里，将从机器学习的视角展示一个机器学习算法实现的基本过程，有别于传统的