手推线性回归算法及代码实现（梯度下降法）

华南师范大学数学科学学院 朱致鹏

假设有样本 X 和标签 Y：

$$X=\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ x_1^{(3)} & x_2^{(3)} & \cdots & x_n^{(3)} \\ & \vdots & & \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} ,Y=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

每行代表一个样本，一共 m 个样本 n 个特征。

我们试图学习这样一个函数：

$$h_\theta(X)=\theta_0 x_0+\theta_1 x_1+\theta_2 x_2+...+\theta_n x_n \qquad (1)$$

其中 $x_0^{(i)}=1$

转化为矩阵形式：

$$\theta=\begin{bmatrix}\theta_0 & \theta_1 & \theta_2 & \cdots & \theta_n\end{bmatrix}$$

$$X=\begin{bmatrix}x_0 & x_1 & x_2 & \cdots & x_n\end{bmatrix}$$

$$h_\theta(X)=X\theta^T=\begin{bmatrix}x_0 & x_1 & x_2 & \cdots & x_n\end{bmatrix}\begin{bmatrix}\theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n\end{bmatrix}$$

定义线性回归的损失函数为（均方误差）：

$\cos t=\sum_{i=1}^{m}(y_i-\hat{y}_i)^2$（$\hat{y}_i$ 表示预测值），表示每一个训练点到直线的竖直距离的平方和。

我们的目标：最小化损失函数求得最佳参数 θ

所以有如下损失函数：

$$J(\theta_0,\theta_1,\theta_2,...,\theta_n)=\frac{1}{2m}\sum_{i=1}^{m}(y_i-h_\theta(x^{(i)}))^2 \qquad (4)$$

通常我们写成 $\frac{1}{2m}\sum_{i=1}^{m}(y_i-h_\theta(x^{(i)}))^2$ 是因为除以 m 可以消除样本数量 m 不同对 $J(\theta)$ 的影响，除以 2 是方便后续计算。

有了损失函数，我们利用梯度下降法求最佳参数 θ：

$$\theta_j=\theta_j-\alpha\frac{\partial}{\partial\theta_j}J(\theta) \qquad (5)$$

$$\frac{\partial}{\partial\theta_j}J(\theta)=-\frac{1}{m}\sum_{i=1}^{m}(y_i-h_\theta(x^{(i)}))x_j^{(i)} \qquad (6)$$

其中，$\frac{\partial h_\theta(x)}{\partial\theta_j}=\frac{\partial(\theta_0+\theta_1x_1+\theta_2x_2+...+\theta_nx_n)}{\partial\theta_j}=x_j \qquad (7)$，这里要注意的是：由于是 $y_i-h_\theta(x)$，所以求导后有一个负号！

因此可以得到：

$$\theta_j=\theta_j-\alpha[-\frac{1}{m}\sum_{i=1}^{m}(y_i-h_\theta(x^{(i)}))x_j^{(i)}] \qquad (8)$$

实际上，矩阵形式为：

$$\underset{R^{(n+1)\times 1}}{\begin{bmatrix}\theta_0\\ \theta_1\\ \theta_2\\ \vdots\\ \theta_n\end{bmatrix}}=\underset{R^{(n+1)\times 1}}{\begin{bmatrix}\theta_0\\ \theta_1\\ \theta_2\\ \vdots\\ \theta_n\end{bmatrix}}-(-\frac{\alpha}{m})\underset{R^{(n+1)\times m}}{\begin{bmatrix}1 & 1 & \cdots & 1\\ x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(m)}\\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(m)}\\ & & \vdots & \\ x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(m)}\end{bmatrix}}\underset{R^{m\times 1}}{\begin{bmatrix}y_1-h_\theta(x^{(1)})\\ y_2-h_\theta(x^{(2)})\\ y_3-h_\theta(x^{(3)})\\ \vdots\\ y_m-h_\theta(x^{(m)})\end{bmatrix}} \qquad (9)$$