所使用的模型：多元线性回归：

$f(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} + b$ 对于输入$\mathbf{x}$的利用$f(\mathbf{x})$来估计输出y

参数确定：

利用最小二乘法来对ω和b进行估计，为了方便计算和讨论，将ω和b变成一个向量形式： $\hat{\boldsymbol{\omega}}$=（$\boldsymbol{\omega}$；$b$），相应的把原来的数据集扩充为$\mathbf{x}$，即：

$$\mathbf{x} = \begin{pmatrix} \mathbf{x_1}^T & 1 \\ \mathbf{x_2}^T & 1 \\ \vdots & \vdots \end{pmatrix}$$

把标记也写成向量形式$\mathbf{y} = (y_1; y_2; \cdots)$，故可得出均方误差为：

$$E_{\hat{\boldsymbol{\omega}}} = (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}})^T (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}})$$

我们试图让均方误差最小，即：

$$\hat{\boldsymbol{\omega}}^* = \arg\min_{\hat{\boldsymbol{\omega}}} (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}})^T (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}})$$

让$E_{\hat{\boldsymbol{\omega}}}$对$\hat{\boldsymbol{\omega}}$求偏导使之为零：

$$\frac{\partial E_{\hat{\boldsymbol{\omega}}}}{\partial \hat{\boldsymbol{\omega}}} = \frac{\partial (\mathbf{y}^\mathrm{T}\mathbf{y} - \hat{\boldsymbol{\omega}}^\mathrm{T}\mathbf{x}^\mathrm{T}\mathbf{y} + \mathbf{x}\hat{\boldsymbol{\omega}}\mathbf{y}^\mathrm{T} + \hat{\boldsymbol{\omega}}^\mathrm{T}\mathbf{x}^\mathrm{T}\mathbf{x}\boldsymbol{\omega})}{\partial \hat{\boldsymbol{\omega}}}$$
$$= \mathbf{0} - \mathbf{x}^\mathrm{T}\mathbf{y} - \mathbf{x}^\mathrm{T}\mathbf{y} + (\mathbf{x}^\mathrm{T}\mathbf{x} + \mathbf{x}^\mathrm{T}\mathbf{x})\hat{\boldsymbol{\omega}} = \mathbf{0}$$

$$\hat{\boldsymbol{\omega}}^* = (\mathbf{x}^\mathrm{T}\mathbf{x})^{-1}\mathbf{x}^\mathrm{T}\mathbf{y}$$

而$\mathbf{x}^\mathrm{T}\mathbf{x}$可能不是满秩矩阵，故采用正则化或梯度下降处理

正则化：

$$E_{\hat{\boldsymbol{\omega}}} = (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}})^T (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}}) + \lambda \|\hat{\boldsymbol{\omega}}\|^2$$

$$\hat{\boldsymbol{\omega}}^{*} = \arg\min_{\hat{\boldsymbol{\omega}}} (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}})^{T}(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}}) + \lambda \|\hat{\boldsymbol{\omega}}\|^{2}$$

$$\frac{\partial E_{\hat{\boldsymbol{\omega}}}}{\partial \hat{\boldsymbol{\omega}}} = \mathbf{0}$$

$$\hat{\boldsymbol{\omega}}^{*} = (\mathbf{x}^{\mathrm{T}}\mathbf{x} + \lambda \mathrm{I})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y}$$

而 $\mathbf{x}^{\mathrm{T}}\mathbf{x} + \lambda \mathrm{I}$ 一定是可逆的，故可以解出解析解。

梯度下降算法：

（1）取初值 $\hat{\boldsymbol{\omega}}^{(0)}$=[0.1;0.1;0.1;0.1]，置 k=0；

（2）计算 $E_{\hat{\boldsymbol{\omega}}}{}^{(k)} = (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}}^{(k)})^{T}(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\omega}}^{(k)})$；

（3）计算梯度 $\nabla E(\hat{\boldsymbol{\omega}}^{(k)})$，并且一维搜索确定 η_k 得到 $\min_{\eta_k \geq 0} E(\hat{\boldsymbol{\omega}}^{(k)} + \eta_k \times (-\nabla E(\hat{\boldsymbol{\omega}}^{(k)}))$

（4）$\hat{\boldsymbol{\omega}}^{(k+1)} = \hat{\boldsymbol{\omega}}^{(k)} + \eta_k \times (-\nabla E(\hat{\boldsymbol{\omega}}^{(k)})$，并计算 $E(\hat{\boldsymbol{\omega}}^{(k+1)})$，当 $\| E(\hat{\boldsymbol{\omega}}^{(k)}) - E(\hat{\boldsymbol{\omega}}^{(k+1)}) \| < \varepsilon$ 时，令 $\hat{\boldsymbol{\omega}}^{*} = \hat{\boldsymbol{\omega}}^{(k)}$，停止迭代

（5）否则置 k=k+1,返回（2）