线性回归

程煦

xcheng8@njust.edu.cn

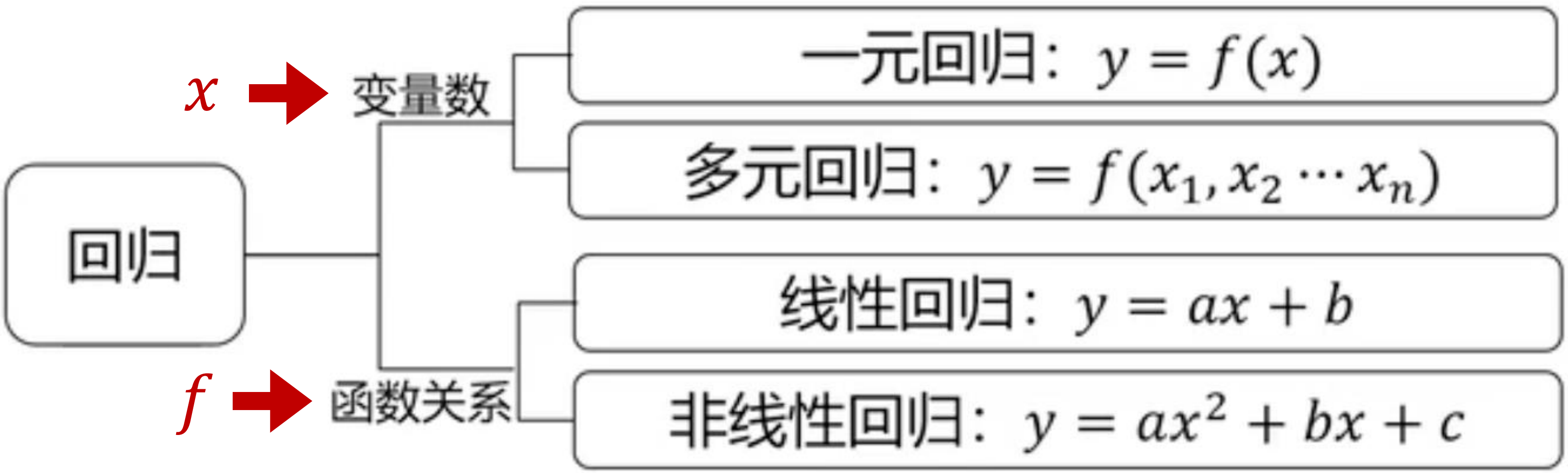
计算机科学与工程学院

回归（regression）

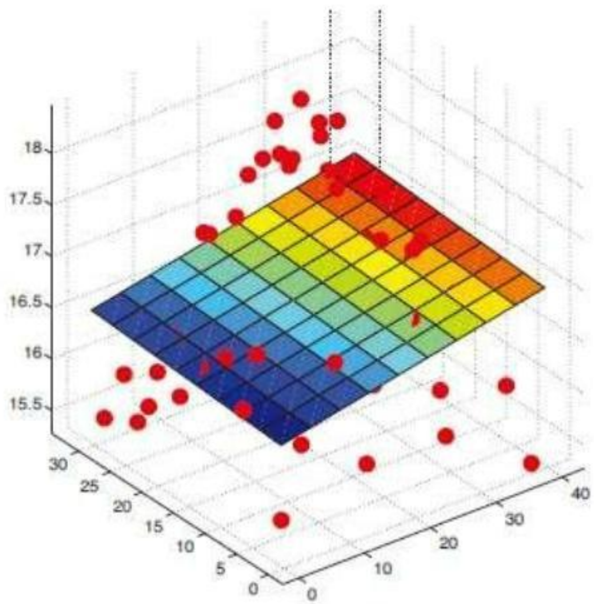
- 问题定义：根据数据，确定输入变量$\boldsymbol{x}^T = [x_1, x_2, \cdots x_{\mathrm{d}}]$与输出$y$之间的定量关系。

- 数学表达式：

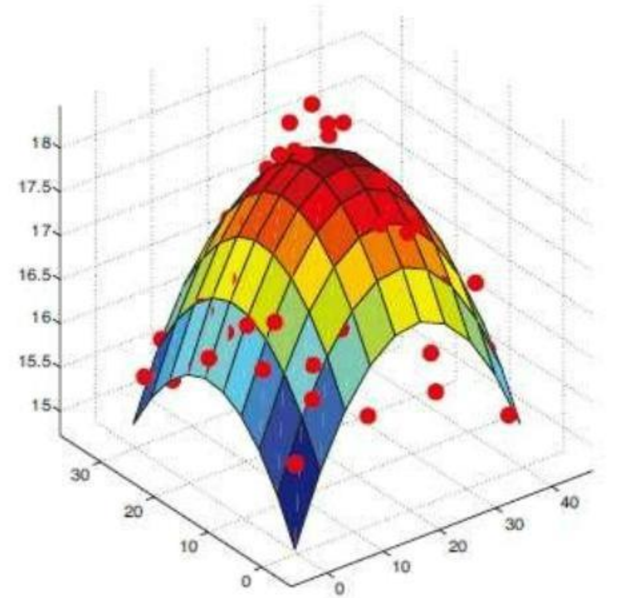
$$y = f(x_1, x_2, \cdots x_{\mathrm{d}})$$



回归（regression）



(a)



(b)

- 考虑建模地理位置对应的温度函数。
  - 平面形式 $\hat{f}(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2.$
  - 二次形式 $\hat{f}(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2.$

线性回归

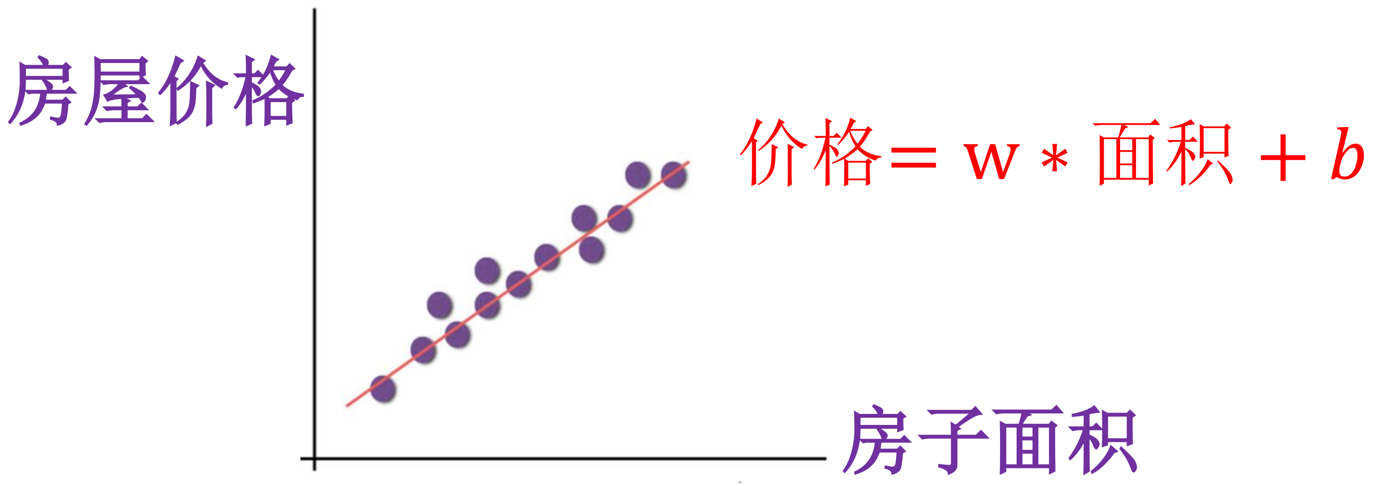
- 定义：输入变量$\boldsymbol{x}^T = [x_1, x_2, \cdots, x_\mathrm{d}]$与输出$y$之存在线性关系

$$y = \mathrm{w}_1 x_1 + \mathrm{w}_2 x_2 + \mathrm{w}_3 x_3 + \cdots + \mathrm{w}_d x_d + b$$

向量形式：
$$y = \boldsymbol{w}^T \boldsymbol{x}$$

$$\boldsymbol{w} = [w_1, w_2, w_3, \cdots, w_d, b],\quad \boldsymbol{x}=[x_1, x_2, x_3, \cdots, x_d, 1]$$

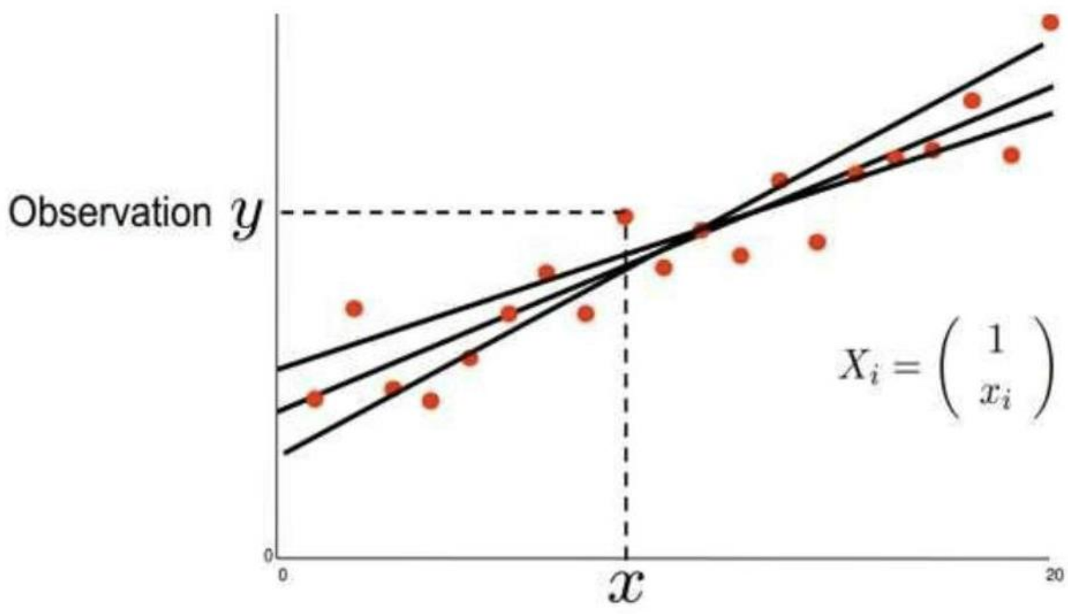
- 举例：



线性回归

- 在众多可能的回归模型中，我们应该选择哪一个?

选择拟合误差小的！



如何定义误差?

均方误差损失

- 给定训练集：$\{(\boldsymbol{x}_i, y_i) | i \in \{1,2,\cdots N\}\}, (\boldsymbol{x}_i)^T = [x_{i1}, x_{i2}, x_{i3}, \cdots, x_{id}]$
- 线性回归假设：

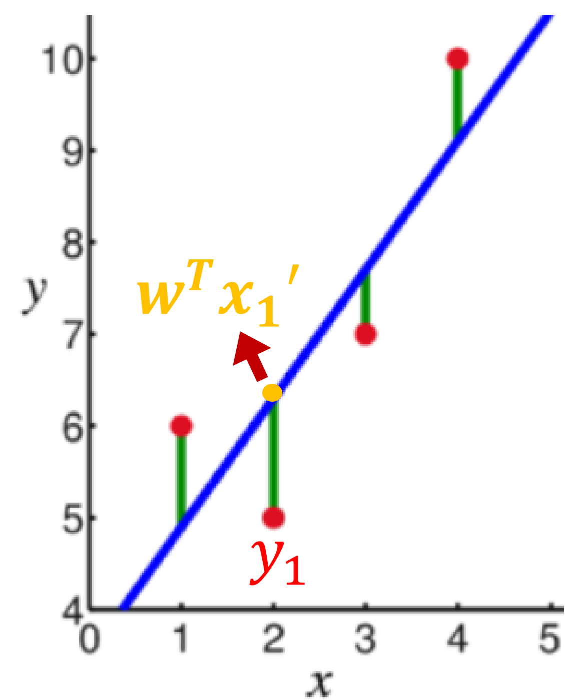
$$\hat{y}_i = \mathrm{w}_1 x_{i1} + \mathrm{w}_2 x_{i2} + \mathrm{w}_3 x_{i3} + \cdots + \mathrm{w}_d x_{id} + b = \boldsymbol{w}^T \boldsymbol{x}_i'$$

$$\boldsymbol{w}^T = [w_1, w_2, w_3, \cdots, w_d, b], \qquad (\boldsymbol{x}_i')^T = [x_{i1}, x_{i2}, \cdots, x_{id}, 1]$$

- 训练集上的经验误差：

$$R(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

$$= \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{w}^T \boldsymbol{x}_i' - y_i)^2$$



如何求解参数 $\boldsymbol{w}$ ？

最小均方误差损失

- 给定训练集：$\{(\boldsymbol{x}_i, y_i) | i \in \{1,2,\cdots N\}\}, (\boldsymbol{x}_i)^T = [x_{i1}, x_{i2}, x_{i3}, \cdots, x_{id}]$
- 线性回归假设：

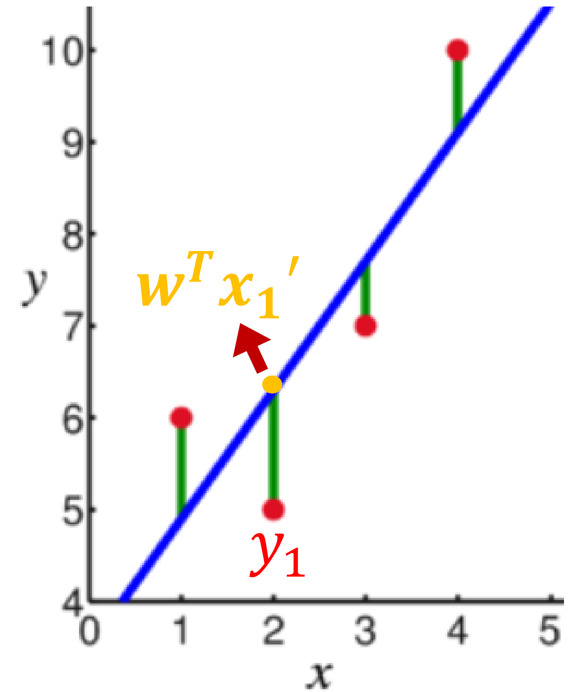
$$\hat{y}_i = \mathrm{w}_1 x_{i1} + \mathrm{w}_2 x_{i2} + \mathrm{w}_3 x_{i3} + \cdots + \mathrm{w}_d x_{id} + b = \boldsymbol{w}^T \boldsymbol{x}_i'$$

$$\boldsymbol{w}^T = [w_1, w_2, w_3, \cdots, w_d, b],\ (\boldsymbol{x}_i')^T = [x_{i1}, x_{i2}, x_{i3}, \cdots, x_{id}, 1]$$

- 最小化经验风险：

$$\boldsymbol{w}^* = \mathrm{argmin}_{\boldsymbol{w}} R(\boldsymbol{w})$$

$$= \mathrm{argmin}_{\boldsymbol{w}} \frac{1}{2} \sum_{i=1}^{N} (\boldsymbol{w}^T \boldsymbol{x}_i' - y_i)^2$$



如何最小化经验风险？

最小均方误差损失求解

- 训练数据： $\boldsymbol{x}_i$列向量

$$\boldsymbol{X}=\begin{bmatrix}(\boldsymbol{x}_1)^T\\(\boldsymbol{x}_2)^T\\\vdots\\(\boldsymbol{x}_N)^T\end{bmatrix}=\begin{bmatrix}x_{11},x_{12},x_{13},\cdots,x_{1d}\\x_{21},x_{22},x_{23},\cdots,x_{2d}\\\vdots\\x_{N1},x_{N2},x_{N3},\cdots,x_{Nd}\end{bmatrix}\qquad \boldsymbol{Y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_N\end{bmatrix}$$

- 线性回归模型：

$$\hat{Y}=\boldsymbol{X}\boldsymbol{w}=\begin{bmatrix}(\boldsymbol{x}_1)^T\cdot\boldsymbol{w}\\(\boldsymbol{x}_2)^T\cdot\boldsymbol{w}\\\vdots\\(\boldsymbol{x}_N)^T\cdot\boldsymbol{w}\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{d}x_{1i}\cdot w_i\\\sum_{i=1}^{d}x_{2i}\cdot w_i\\\vdots\\\sum_{i=1}^{d}x_{Ni}\cdot w_i\end{bmatrix}\qquad \boldsymbol{w}=\begin{bmatrix}w_1\\w_2\\\vdots\\w_d\end{bmatrix}$$

最小均方误差损失求解

- 预测误差

$$\widehat{\boldsymbol{Y}} - \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y} = \begin{bmatrix} (\boldsymbol{x}_1)^T \cdot \boldsymbol{w} \\ (\boldsymbol{x}_2)^T \cdot \boldsymbol{w} \\ \vdots \\ (\boldsymbol{x}_N)^T \cdot \boldsymbol{w} \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} (\boldsymbol{x}_1)^T \cdot \boldsymbol{w} - y_1 \\ (\boldsymbol{x}_2)^T \cdot \boldsymbol{w} - y_2 \\ \vdots \\ (\boldsymbol{x}_N)^T \cdot \boldsymbol{w} - y_N \end{bmatrix}$$

- 均方误差损失

$$R(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}\left((\boldsymbol{x}_i)^T \cdot \boldsymbol{w} - y_i\right)^2$$

$$= \frac{1}{2}(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y})$$

标量 $R(\boldsymbol{w}) \in \mathbb{R}$

最小均方误差损失闭式解

- 目标

$$\min_{\boldsymbol{w}} \frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})$$

- 计算梯度

$$\nabla_{\boldsymbol{w}}R(\boldsymbol{w}) = \nabla_{\boldsymbol{w}}\frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})$$

$$= \frac{1}{2}\nabla_{\boldsymbol{w}}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right)$$

$$= \frac{1}{2}\nabla_{\boldsymbol{w}}\,\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$$

$$\boxed{\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}}$$

$$= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$$

$$\boxed{\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B}+\mathbf{B}^T)\mathbf{x}}$$

最小均方误差损失闭式解

- 目标

$$\min_{\boldsymbol{w}} \frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^{T}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})$$

- 令梯度$\nabla_{\boldsymbol{w}} R(\boldsymbol{w}) = \boldsymbol{0}$, we obtain the closed-form solution

$$令\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$$

$$\Rightarrow \quad \boldsymbol{w}^{*}=\boxed{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$$

1. 难计算
2. 不可逆

Ridge regression（岭回归）

- 预测误差

$$\widehat{\boldsymbol{Y}} - \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y} = \begin{bmatrix} (\boldsymbol{x}_1)^T \cdot \boldsymbol{w} \\ (\boldsymbol{x}_2)^T \cdot \boldsymbol{w} \\ \vdots \\ (\boldsymbol{x}_N)^T \cdot \boldsymbol{w} \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} (\boldsymbol{x}_1)^T \cdot \boldsymbol{w} - y_1 \\ (\boldsymbol{x}_2)^T \cdot \boldsymbol{w} - y_2 \\ \vdots \\ (\boldsymbol{x}_N)^T \cdot \boldsymbol{w} - y_N \end{bmatrix}$$

- Ridge regression

$$R(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}\left((\boldsymbol{x}_i)^T \cdot \boldsymbol{w} - y_i\right)^2 + \frac{1}{2}\lambda\sum_{i=1}^{d}(w_i)^2$$

$$= \frac{1}{2}(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y})^T\,(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{Y}) + \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2$$

标量 $R(\boldsymbol{w}) \in \mathbb{R}$

Ridge regression闭式解

- 目标

$$\min_{\boldsymbol{w}} \frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})+\frac{\lambda}{2}\|\boldsymbol{w}\|_2^2$$

- 计算梯度

$$\nabla_{\boldsymbol{w}} R(\boldsymbol{w}) = \nabla_{\boldsymbol{w}} \frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y}) + \boxed{\nabla_{\boldsymbol{w}} \frac{\lambda}{2}(\boldsymbol{w})^T\boldsymbol{w}}$$

标量对向量的求导

在矩阵微积分中，**标量对向量求导**的基本公式是：

$$\nabla_{\mathbf{w}} f(\mathbf{w}) = \left[\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}, \ldots, \frac{\partial f}{\partial w_d}\right]^T$$

即，标量函数 $f(\mathbf{w})$ 对向量 $\mathbf{w}$ 求导的结果是一个列向量，表示每个分量的偏导数。

标量对向量的求导（回顾）

$$\nabla_{\mathbf{w}} f = \nabla_{\mathbf{w}}(\mathbf{w}^T\mathbf{w}) = 2\mathbf{w}.$$

其中，$\mathbf{w}$ 是一个 $d \times 1$ 的列向量，即：

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}.$$

目标函数是：

$$f(\mathbf{w}) = \mathbf{w}^T\mathbf{w}.$$

展开成分量形式：

$$f(\mathbf{w}) = w_1^2 + w_2^2 + \cdots + w_d^2.$$

对每个分量 w_i 计算偏导数：

$$\frac{\partial f}{\partial w_i} = \frac{\partial}{\partial w_i}\sum_{j=1}^{d} w_j^2 = 2w_i.$$

因此，梯度向量为：

$$\nabla_{\mathbf{w}} f = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_d} \end{bmatrix} = \begin{bmatrix} 2w_1 \\ 2w_2 \\ \vdots \\ 2w_d \end{bmatrix} = 2\mathbf{w}.$$

Ridge regression闭式解

- 目标

$$\min_{\boldsymbol{w}} \frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})+\frac{\lambda}{2}\|\boldsymbol{w}\|_2^2$$

- 计算梯度

$$\nabla_{\boldsymbol{w}} R(\boldsymbol{w})=\nabla_{\boldsymbol{w}}\frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})+\boxed{\nabla_{\boldsymbol{w}}\frac{\lambda}{2}(\boldsymbol{w})^T\boldsymbol{w}}$$

标量对向量的求导

$$=\frac{1}{2}\nabla_{\boldsymbol{w}}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right)+\lambda\boldsymbol{w}$$

$$=\frac{1}{2}\nabla_{\boldsymbol{w}}\,\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}+\lambda\boldsymbol{w}$$

$$=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}+\lambda\boldsymbol{w}$$

Ridge regression闭式解

- 目标

$$\min_{\boldsymbol{w}} \ \frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})^{T}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{Y})+\frac{\lambda}{2}\|\boldsymbol{w}\|_2^2$$

- 令梯度$\nabla_{\boldsymbol{w}} R(\boldsymbol{w})=\mathbf{0}$, we obtain the closed-form solution

$$令\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}+\lambda\boldsymbol{w}=\mathbf{0}$$

$$\rightarrow \quad \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}+\lambda\boldsymbol{w}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$$

$$\rightarrow \quad \boldsymbol{w}^{*}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}+\lambda\boldsymbol{I})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$$

$\boldsymbol{I}$： $d\times d$的单位阵

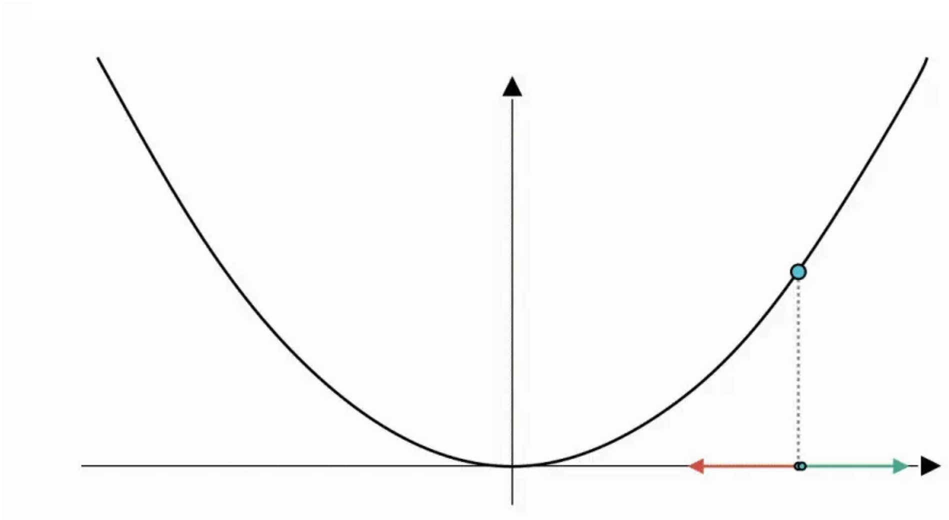
- 当 $\lambda=0$ 时，Ridge 回归退化为普通的最小二乘回归（OLS）。
- 当 λ 较大时，模型倾向于更小的权重，从而更偏向简单模型。

数值优化的梯度下降法（Gradient Descend, GD）

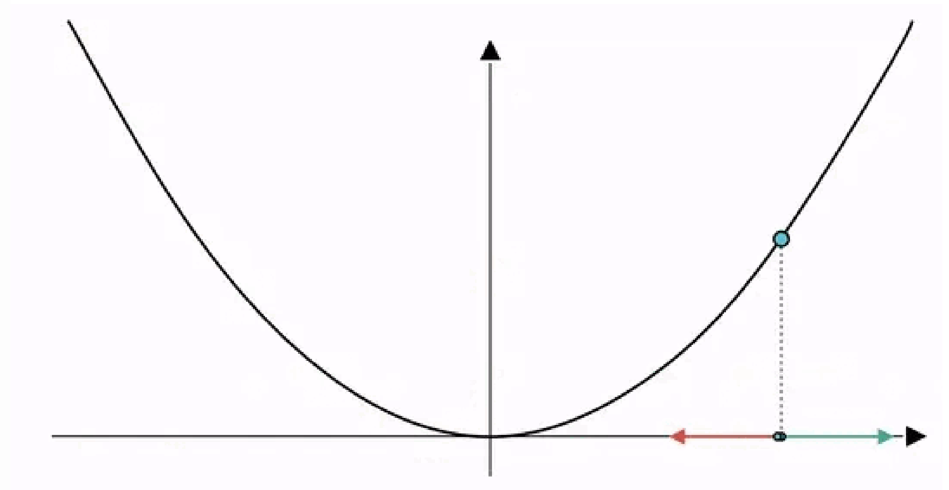
- 梯度下降法是寻找函数$R(\boldsymbol{w})$最小值的一阶优化迭代算法
- 主要思想：梯度反方向是数值下降最快的方向

1. 梯度的定义：

梯度是一个向量，它指向目标函数值增加最快的方向，且其大小表示沿着该方向的变化速率。具体来说，如果目标函数 $R(\mathbf{w})$ 在点 $\mathbf{w}$ 处有梯度 $\nabla_{\mathbf{w}} R(\mathbf{w})$，那么它指向目标函数值上升最快的方向。



梯度方向



梯度反方向

函数值越来越小

数值优化的梯度下降法（Gradient Descend, GD）

- 优化过程：通过向函数在当前点对应的梯度（或近似梯度，通过目标函数的局部变化来近似）的反方向的规定步长距离点进行迭代搜索，直至收敛。

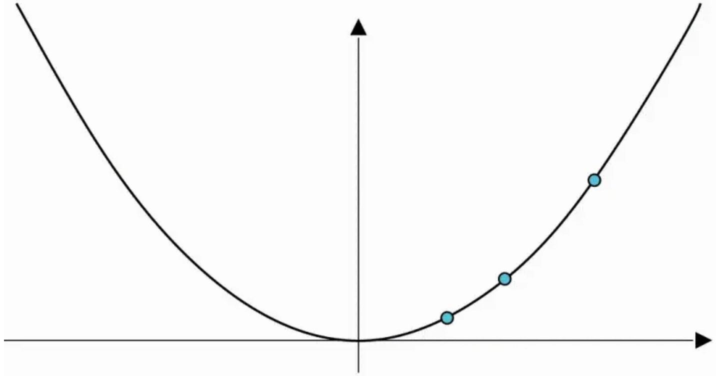
①从初始位置开始：　初始参数$\boldsymbol{w}_t, t=0$；

②计算当前位置梯度: $\nabla_{\boldsymbol{w}} R(\boldsymbol{w})|_{\boldsymbol{w}=\boldsymbol{w}_t}$；

③沿梯度反方向移动到下一个位置 $t+1$:

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \alpha \nabla_{\boldsymbol{w}} R(\boldsymbol{w})|_{\boldsymbol{w}=\boldsymbol{w}_t},\quad \alpha \text{ 是步长}$$

④ 重复②和③直到收敛，得到$\boldsymbol{w}^*$。



最小均方误差损失优化

■ 线性回归的梯度下降

① 计算梯度：列向量 $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{d\times 1}}$, 列向量 $\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{d\times 1}}$

$$\frac{\partial R(\boldsymbol{w})}{\partial \boldsymbol{w}} = \frac{1}{2}\frac{\partial}{\partial \boldsymbol{w}}\sum_{i=1}^{N}((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i)^2$$

$$= \frac{1}{2} \cdot 2\sum_{i=1}^{N}((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i) \cdot \frac{\partial}{\partial \boldsymbol{w}}((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i)$$

$$= \sum_{i=1}^{N}((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i)\boxed{\frac{\partial}{\partial \boldsymbol{w}}((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i)}$$

标量对列向量的求导，求导结果与$\boldsymbol{w}$相同维度的列向量

$$= \sum_{i=1}^{N}\boxed{((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i)\boldsymbol{x_i}}$$

“误差× 输入”

最小均方误差损失优化

■ 线性回归的梯度下降

① 计算梯度：

$$\frac{\partial R(\boldsymbol{w})}{\partial \boldsymbol{w}} = \sum_{i=1}^{N} ((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i)\boldsymbol{x_i}$$

② 重复沿着梯度反方向更新参数$\boldsymbol{w}$，直至收敛：

$$\boldsymbol{w} := \boldsymbol{w} - \alpha \frac{\partial}{\partial \boldsymbol{w}} R(\boldsymbol{w})$$

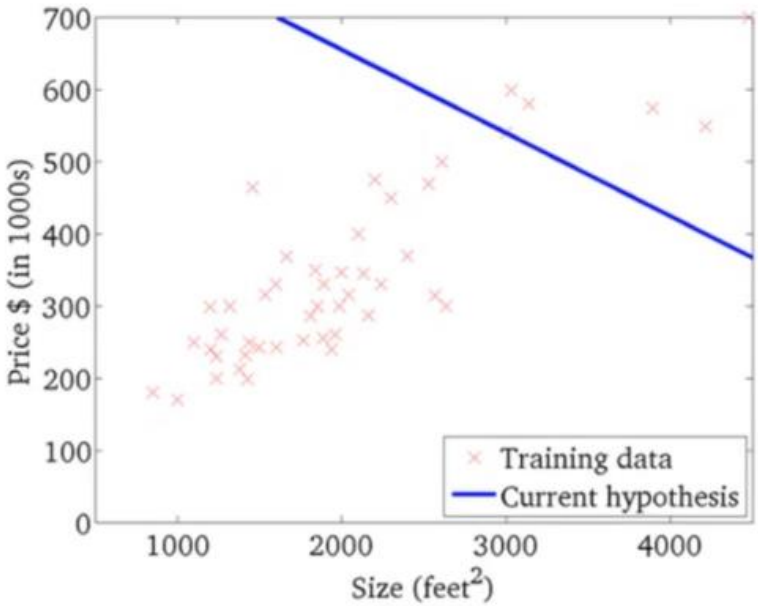
$$= \boldsymbol{w} - \alpha \sum_{i=1}^{N} ((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i)\boldsymbol{x_i}$$

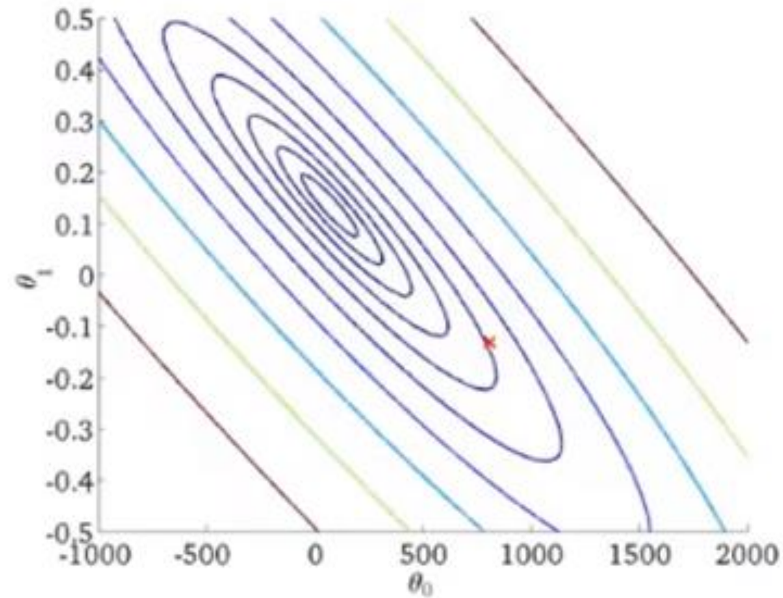
最小均方误差损失优化

■ 线性回归的梯度下降

求解目标：$y = [w_{1,} w_2]^T x$

经验风险：$R(w_{1,}, w_2)$
(function of the parameter w_1, w_2)



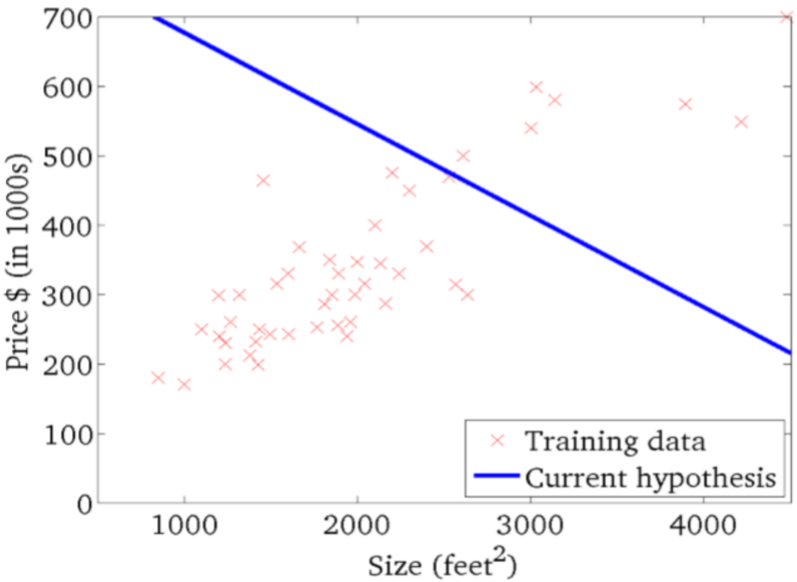


Step 0

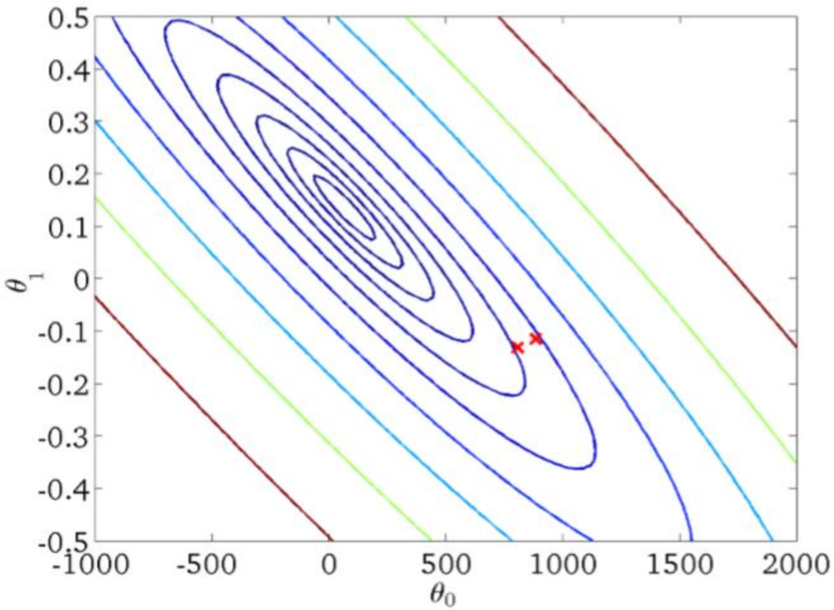
最小均方误差损失优化

■ 线性回归的梯度下降

求解目标： $y = [\mathrm{w}_{1,}\mathrm{w}_2]^T x$



经验风险 ： $R(\mathrm{w}_{1,}, \mathrm{w}_2)$
(function of the parameter $\mathrm{w}_1, \mathrm{w}_2$)



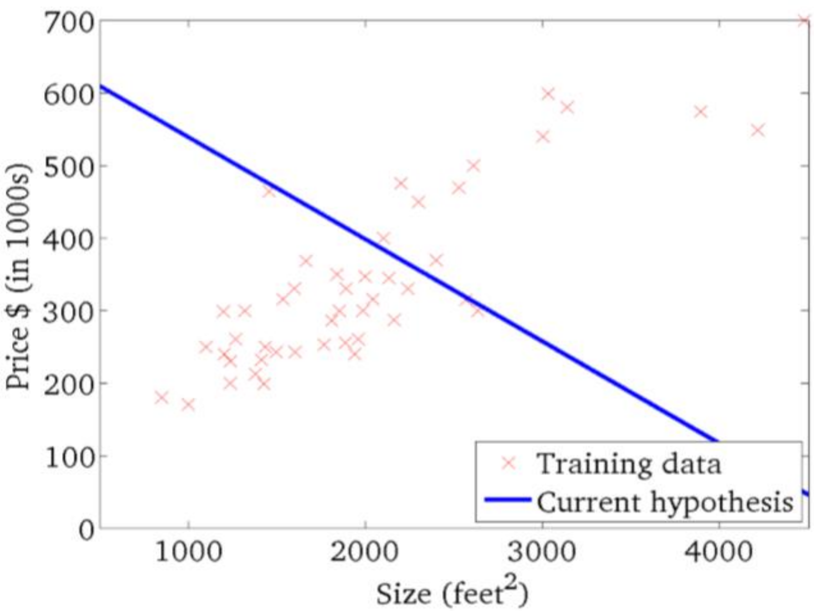
Step 1

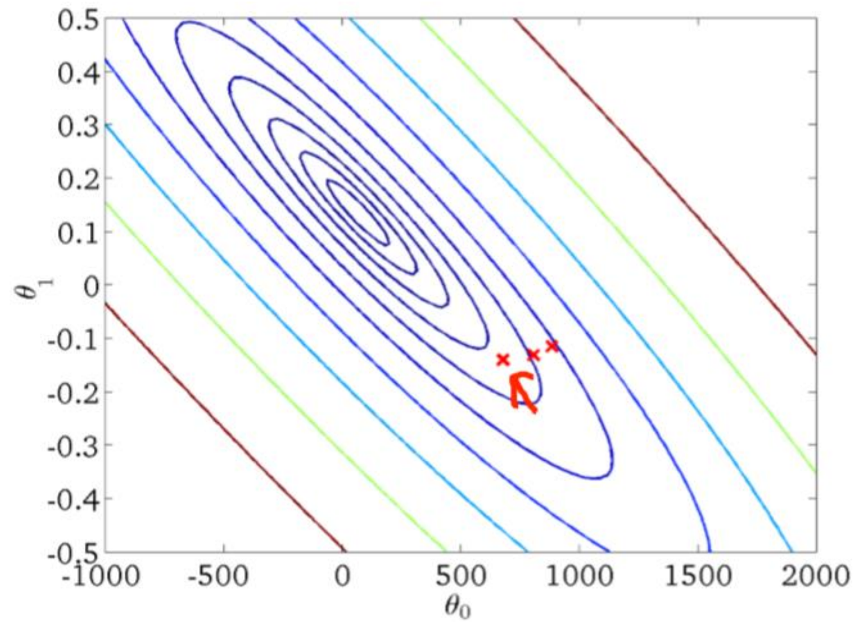
最小均方误差损失优化

■ 线性回归的梯度下降

求解目标：$y = [\mathrm{w}_1, \mathrm{w}_2]^T x$

经验风险：$R(\mathrm{w}_1, \mathrm{w}_2)$
(function of the parameter $\mathrm{w}_1, \mathrm{w}_2$)



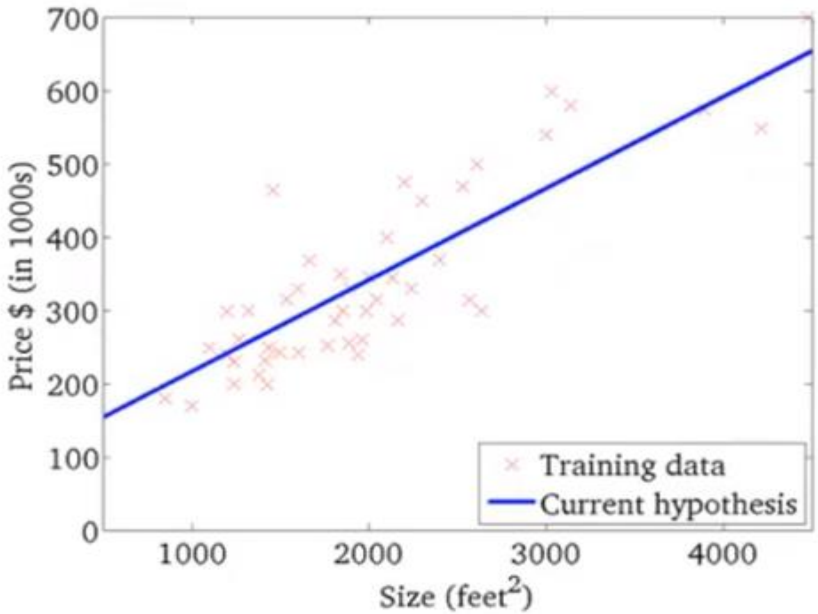


Step 2

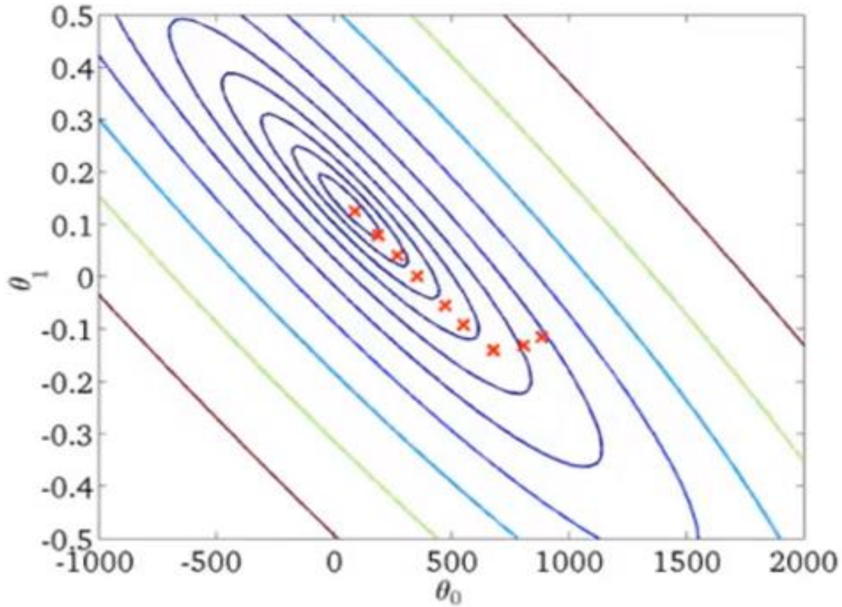
最小均方误差损失优化

■ 线性回归的梯度下降

求解目标：$y = [\mathrm{w}_1,\mathrm{w}_2]^T x$



经验风险：$R(\mathrm{w}_1,\mathrm{w}_2)$
(function of the parameter $\mathrm{w}_1, \mathrm{w}_2$)



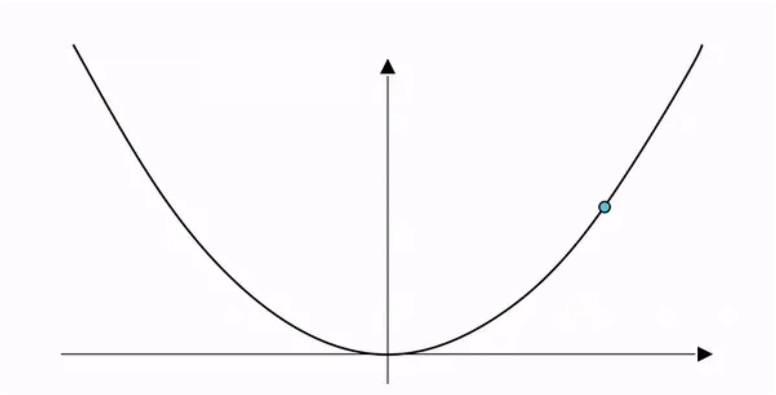
Final Step

最小均方误差损失优化

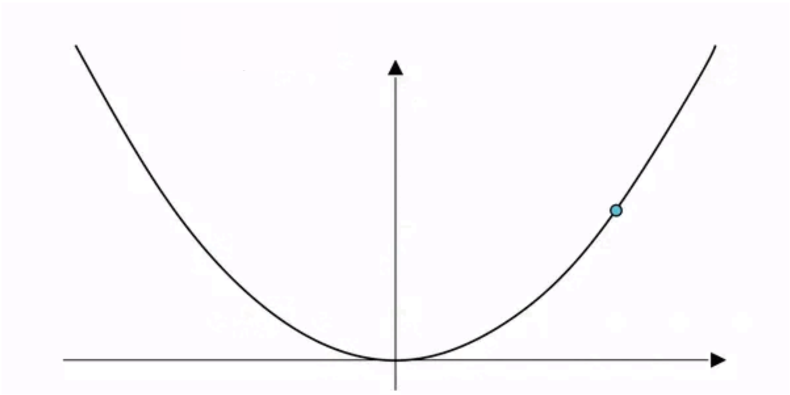
- 线性回归的梯度下降

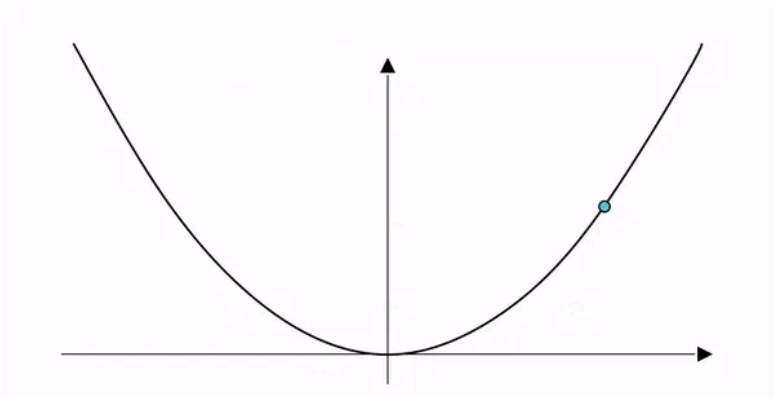
  - 步长α的选择　$\boldsymbol{w} := \boldsymbol{w} - \alpha \frac{\partial}{\partial \boldsymbol{w}} R(\boldsymbol{w})$

步长太大：跳过最小值，难以收敛



步长太小：迭代次数增多，收敛速度变慢





Tips: 选择合适的、较小的步长

最小均方误差损失优化

■ 梯度下降种类

① 批量梯度下降（Batch Gradient Descent，BGD）

- 在每一次迭代中使用全部训练样本来计算梯度、更新参数
- 计算时间复杂度高，适用于小数据集的情况

$$\boldsymbol{w} := \boldsymbol{w} - \alpha \sum_{i=1}^{N} \left((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i \right) \boldsymbol{x_i}$$

② 随机梯度下降（Stochastic Gradient Descent，SGD）

- 每次只使用一个样本计算梯度，并更新模型参数
- 更新过程非常噪声，可能导致目标函数的波动较大，收敛过程不如批量梯度下降稳定，也不一定收敛到全局最小值

$$\boldsymbol{w} := \boldsymbol{w} - \alpha \left((\boldsymbol{x_i})^T \cdot \boldsymbol{w} - y_i \right) \boldsymbol{x_i}$$

最小均方误差损失优化

■ 梯度下降种类

③ 小批量梯度下降（Mini-Batch Gradient Descent）

- 每次迭代中，使用一小部分样本（称为 mini-batch）来计算梯度并更新参数
- 通过减小每次更新的计算量，提高了效率，同时通过使用小批量样本来减少噪声，使得梯度更新较为稳定。

$$\boldsymbol{w} := \boldsymbol{w} - \alpha \sum_{i=1}^{N'} ((\boldsymbol{x}_i)^T \cdot \boldsymbol{w} - y_i)\boldsymbol{x}_i$$

对数线性回归

- 线性模型虽简单,却有丰富的变化。例如对于样本(x, y)，当我们希望线性模型的预测值逼近真实值y时,就得到了线性回归模型。为便于观察,我们把线性回归模型简写为

$$
\begin{aligned}
\hat{y} &= \mathrm{w}_1 x_1 + \mathrm{w}_2 x_2 + \mathrm{w}_3 x_3 + \cdots + \mathrm{w}_d x_d + b \\
&= \boldsymbol{w}^T \boldsymbol{x}
\end{aligned}
$$

$$\boldsymbol{w} = [w_1, w_2, w_3, \cdots, w_d, b], \quad \boldsymbol{x}=[x_1, x_2, x_3, \cdots, x_d, 1]$$

- 可否令模型预测值逼近y的衍生物呢？譬如说，假设我们认为样本所对应的输出$\hat{y}$是在指数尺度上变化？

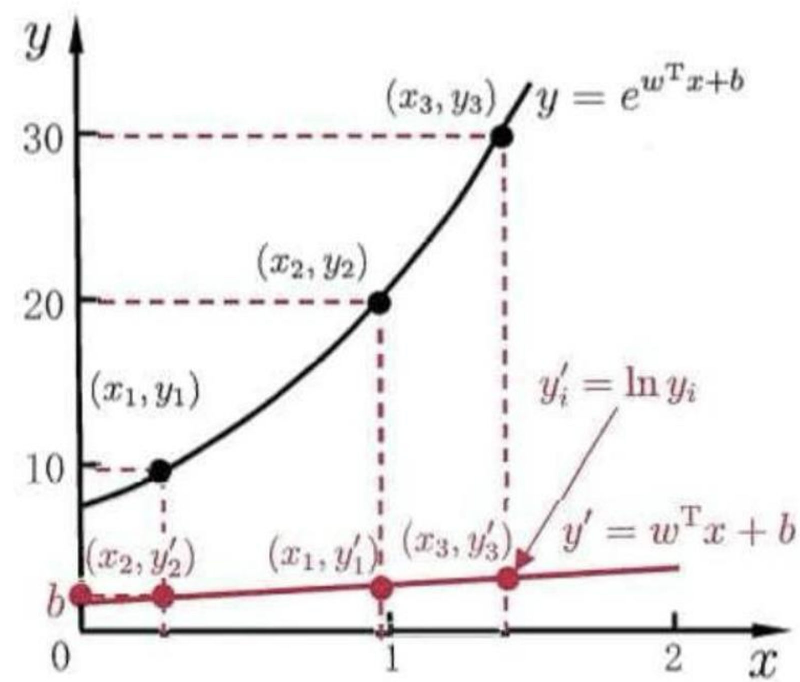
对数线性回归

- 形式：

$$\hat{y} = e^{w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_d x_d + b} = e^{\boldsymbol{w}^T \boldsymbol{x}}$$

$$\Rightarrow \ln \hat{y} = \boldsymbol{w}^T \boldsymbol{x}$$

- 将输出y的对数作为线性模型逼近的目标，实质上求取输入空间到输出空间的非线性映射。




Thank You!