多元线性回归

求解权重 $\hat{w}$ 的公式推导

推导思路: 由最小二乘法导出损失函数 $E_{\hat{w}}$

↓

证明损失函数 $E_{\hat{w}}$ 是关于 $\hat{w}$ 的凸函数

↓

对损失函数 $E_{\hat{w}}$ 关于 $\hat{w}$ 求一阶导数

↓

令一阶导数等于 0 解出 $\hat{w}^*$

将 w 和 b 组合成 $\hat{w}$: $f(x_i) = w^T x_i + b$

多元线性回归: $f(x_i) = (w_1, w_2, \cdots w_d) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{id} \end{pmatrix} + b$

$f(x_i) = w_1 x_1 + w_2 x_2 + \cdots + w_d x_{id} + b$

$f(x_i) = w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} + w_{d+1} \cdot 1$

$f(x_i) = (w_1 \quad w_2 \quad \cdots \quad w_d \quad \boxed{w_{d+1}}) \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \\ \boxed{1} \end{pmatrix}$

$f(\hat{x}_i) = \hat{w}^T \hat{x}_i$

③

由最小二乘法导出损失函数 $E_{\hat{w}}$: $\quad f(\hat{x}_i)=\hat{w}^T x_i$

$$E_{\hat{w}}=\sum_{i=1}^{m}(y_i-f(\hat{x}_i))^2$$

$$=\sum_{i=1}^{m}(y_i-\hat{w}^T\hat{x}_i)^2$$

$$X=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{pmatrix}=\begin{pmatrix} x_1^T & 1 \\ x_2^T & 1 \\ \vdots & \vdots \\ x_m^T & 1 \end{pmatrix}=\begin{pmatrix} \hat{x}_1^T \\ \hat{x}_2^T \\ \vdots \\ \hat{x}_m^T \end{pmatrix}$$

$$y=(y_1,y_2,\cdots y_m)^T$$

$$E_{\hat{w}}=\sum_{i=1}^{m}(y_i-\hat{w}^T\hat{x}_i)^2$$

$$=(y_1-\hat{w}^T\hat{x}_1)^2+(y_2-\hat{w}^T\hat{x}_2)^2+\cdots+(y_m-\hat{w}^T\hat{x}_m)^2$$

$$E_{\hat{w}}=(y_1-\hat{w}^T\hat{x}_1 \quad y_2-\hat{w}^T\hat{x}_2 \quad \cdots \quad y_d-\hat{w}^T\hat{x}_d)\begin{pmatrix} y_1-\hat{w}^T\hat{x}_1 \\ y_2-\hat{w}^T\hat{x}_2 \\ \vdots \\ y_d-\hat{w}^T\hat{x}_d \end{pmatrix}=(y-X\hat{w})^T(y-X\hat{w})$$

即3.9

又因为

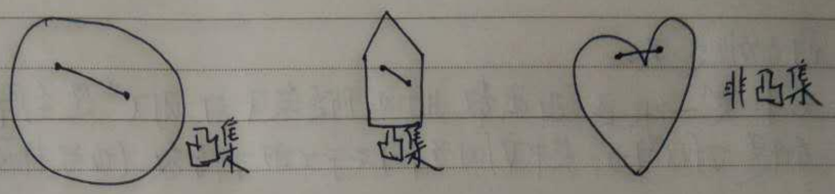
$$\begin{pmatrix} y_1-\hat{w}^T\hat{x}_1 \\ y_2-\hat{w}^T\hat{x}_2 \\ \vdots \\ y_d-\hat{w}^T\hat{x}_d \end{pmatrix}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix}-\begin{pmatrix} \hat{w}^T\hat{x}_1 \\ \hat{w}^T\hat{x}_2 \\ \vdots \\ \hat{w}^T\hat{x}_d \end{pmatrix}$$

$$=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix}-\begin{pmatrix} \hat{x}_1^T\hat{w} \\ \hat{x}_2^T\hat{w} \\ \vdots \\ \hat{x}_d^T\hat{w} \end{pmatrix}=y-X\hat{w}$$

$$\begin{pmatrix} \hat{x}_1^T\hat{w} \\ \hat{x}_2^T\hat{w} \\ \vdots \\ \hat{x}_d^T\hat{w} \end{pmatrix}=\begin{pmatrix} \hat{x}_1^T \\ \hat{x}_2^T \\ \vdots \\ \hat{x}_m^T \end{pmatrix}\cdot\hat{w}=X\cdot\hat{w}$$

凸集定义：设集合 $D \in R^n$，如果对任意的 $x, y \in D$ 与任意的 $\alpha \in [0,1]$，有 $\alpha x + (1-\alpha) y \in D$，则称集合 D 是凸集

凸集的几何意义是：若两个点属于此集合，则这两点连线上的任意一点均属于此集合



梯度定义：设 n 元函数 $f(x)$ 对自变量 $x=(x_1, x_2, \cdots x_n)^T$ 的各分量 x_i 的偏导数 $\frac{\partial f(x)}{\partial x_i}$ $(i=1,2,\cdots,n)$ 都存在，则称函数 $f(x)$ 在 x 处一阶可导，并称向量

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

为函数 $f(x)$ 在 x 处的一阶导数或梯度，记为 $\nabla f(x)$
列向量.

Hessian（海塞）矩阵定义：设 n 元函数 $f(x)$ 对自变量 $x=(x_1, x_2, \cdots x_n)^T$ 的各分量 x_i 的二阶偏导数 $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ $(i=1,2,\cdots,n;\ j=1,2,\cdots,n)$ 都存在，则称函数 $f(x)$ 在点 x 处二阶可导，并称矩阵

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

为 $f(x)$ 在 x 处的二阶导数或 Hessian 矩阵，记为 $\nabla^2 f(x)$，若 $f(x)$ 对 x 各变元的所有二阶偏导数都连续，则 $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$ 此时 $\nabla^2 f(x)$ 为对称矩阵.

④

多元实值函数凹凸性判定定理：

设 $D \subset R^n$ 是非空开凸集，$f: D \subset R^n \to R$，且 $f(x)$ 在 D 上二阶连续可微，如果 $f(x)$ 的 Hessian 矩阵 $\nabla^2 f(x)$ 在 D 上是正定的，则 $f(x)$ 是 D 上的严格凸函数。

凸充分性定理：

若 $f: R^n \to R$ 是凸函数，且 $f(x)$ 一阶连续可微，则 x^* 是全局解的充分必要条件是 $\nabla f(x^*)=0$，其中 $\nabla f(x)$ 为 $f(x)$ 关于 x 的一阶导数（也称梯度）

证明损失函数 $E_{\hat{w}}$ 是关于 $\hat{w}$ 的凸函数：

$$\frac{\partial E_{\hat{w}}}{\partial \hat{w}} = \frac{\partial}{\partial \hat{w}}\left[(y - X\hat{w})^T (y - X\hat{w})\right]$$

$$= \frac{\partial}{\partial \hat{w}}\left[(y^T - \hat{w}^T X^T)(y - X\hat{w})\right]$$

$$= \frac{\partial}{\partial \hat{w}}\left[y^T y - y^T X\hat{w} - \hat{w}^T X^T y + \hat{w}^T X^T X\hat{w}\right]$$

$$= \frac{\partial}{\partial \hat{w}}\left[-y^T X\hat{w} - \hat{w}^T X^T y + \hat{w}^T X^T X\hat{w}\right]$$

[标量-向量]的矩阵微分公式为：

$$\frac{\partial y}{\partial x} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix} \qquad \frac{\partial y}{\partial x} = \left(\frac{\partial y}{\partial x_1}\ \frac{\partial y}{\partial x_2}\ \cdots\ \frac{\partial y}{\partial x_n}\right)$$

（分母布局）[默认采用]　　　　（分子布局）

其中，$x = (x_1, x_2, \cdots x_n)^T$ 为 n 维列向量，y 为 x 的 n 元标量函数。

由[标量—向量]的矩阵微分公式可推得:

$$\frac{\partial x^Ta}{\partial x} = \frac{\partial a^Tx}{\partial x} = \begin{pmatrix} \frac{\partial(a_1x_1+a_2x_2+\cdots+a_nx_n)}{\partial x_1} \\ \frac{\partial(a_1x_1+a_2x_2+\cdots+a_nx_n)}{\partial x_2} \\ \vdots \\ \frac{\partial(a_1x_1+a_2x_2+\cdots+a_nx_n)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a$$

同理可推得: $\frac{\partial x^TBx}{\partial x} = (B+B^T)x$

$$\frac{\partial E_{\hat{w}}}{\partial \hat{w}} = \frac{\partial}{\partial \hat{w}}\left[-y^TX\hat{w} - \hat{w}^TX^Ty + \hat{w}^TX^TX\hat{w}\right]$$

$$= \left[-\frac{\partial y^TX\hat{w}}{\partial \hat{w}} - \frac{\partial \hat{w}^TX^Ty}{\partial \hat{w}} + \frac{\partial \hat{w}^TX^TX\hat{w}}{\partial \hat{w}}\right]$$

由矩阵微分公式: $\frac{\partial x^Ta}{\partial x} = \frac{\partial a^Tx}{\partial x} = a$

$\frac{\partial x^TBx}{\partial x} = (B+B^T)x$ 可得:

$$\frac{\partial E_{\hat{w}}}{\partial \hat{w}} = -X^Ty - X^Ty + (X^TX + X^TX)\hat{w}$$

$$= 2X^T(X\hat{w} - y) \quad \text{即 } 3.10$$

$$\frac{\partial^2 E_{\hat{w}}}{\partial \hat{w}\partial \hat{w}^T} = \frac{\partial}{\partial \hat{w}}\left(\frac{\partial E_{\hat{w}}}{\partial \hat{w}}\right)$$

$$= \frac{\partial}{\partial \hat{w}}\left[2X^T(X\hat{w} - y)\right]$$

$$= \frac{\partial}{\partial \hat{w}}\left(2X^TX\hat{w} - 2X^Ty\right)$$

$= 2X^TX$ 此即为Hessian矩阵

⑥

对损失函数 $E_{\hat{w}}$ 关于 $\hat{w}$ 求一阶导数:

$$\frac{\partial E_{\hat{w}}}{\partial \hat{w}} = 2X^T(X\hat{w}-y)$$

令一阶导数等于0解出 $\hat{w}^*$.

$$\frac{\partial E_{\hat{w}}}{\partial \hat{w}} = 2X^T(X\hat{w}-y)=0$$

$$2X^TX\hat{w}-2X^Ty=0$$

$$2X^TX\hat{w}=2X^Ty$$

$$\hat{w}^*=(X^TX)^{-1}X^Ty$$ 即为3.11