线性回归

2021-09-04

1. 一元线性回归

1.1. 模型定义

模型定义如下所示:

$$f(x) = \hat{y} = wx + b \\ \text{其中,} w\text{和}b\text{为模型要学习的参数,} x\text{为样本值} \tag{30}$$

根据最小二乘法，模型的损失函数为:

$$\begin{aligned} \mathbf{E(w,b)} &= \sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{m}(y_i - (wx_i + b))^2 \end{aligned} \\ \text{其中,} y_i\text{为真实值,} \hat{y}_i\text{为预测值,} m\text{为样本个数} \tag{31}$$

模型优化目标：找到合适的w, b，使得$\mathbf{E(w,b)}$的值最小

$$w, b = \arg\min_{w,b} \mathbf{E(w,b)} \tag{32}$$

1.2. 损失函数凹凸性证明

证明损失函数$\mathbf{E(w,b)}$是关于w和b的凸函数。

1. 求$A = f''_{xx}(x, y)$

$$\begin{aligned} \frac{\partial E(w,b)}{\partial w} &= \frac{\partial}{\partial w}\left[\sum_{i=1}^{m}(y_i - wx_i - b)^2\right] \\ &= \sum_{i=1}^{m}\frac{\partial}{\partial w}(y_i - wx_i - b)^2 \\ &= \sum_{i=1}^{m} 2 \cdot (y_i - wx_i - b) \cdot (-x_i) \\ &= 2\left(w\sum_{i=1}^{m}x_i^2 - \sum_{i=1}^{m}(y_i - b)x_i\right) \end{aligned} \tag{33}$$

$$\begin{aligned} \frac{\partial^2 E(w,b)}{\partial w^2} &= \frac{\partial}{\partial w}\left(\frac{\partial E(w,b)}{\partial w}\right) \\ &= \frac{\partial}{\partial w}\left[2\left(w\sum_{i=1}^{m}x_i^2 - \sum_{i=1}^{m}(y_i - b)x_i\right)\right] \end{aligned} \tag{34}$$

$$
\begin{aligned}
&= \frac{\partial}{\partial w}\left[2w\sum_{i=1}^{m} x_i^2\right] \\
&= 2\sum_{i=1}^{m} x_i^2
\end{aligned}
$$

所以，$A = 2\sum_{i=1}^{m} x_i^2$

2. 求$B = f''_{xy}(x,y)$

$$
\begin{aligned}
\frac{\partial^2 E(w,b)}{\partial w \partial b} &= \frac{\partial}{\partial b}\left(\frac{\partial E(w,b)}{\partial w}\right) \\
&= \frac{\partial}{\partial b}\left[2\left(w\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m}(y_i - b)x_i\right)\right] \\
&= \frac{\partial}{\partial b}\left[-2\sum_{i=1}^{m}(y_i - b)x_i\right] \\
&= \frac{\partial}{\partial b}\left(-2\sum_{i=1}^{m} y_i x_i + 2\sum_{i=1}^{m} b x_i\right) \\
&= \frac{\partial}{\partial b}\left(2\sum_{i=1}^{m} b x_i\right) \\
&= 2\sum_{i=1}^{m} x_i
\end{aligned}
\tag{35}
$$

所以，$B = 2\sum_{i=1}^{m} x_i$

3. 求$C = f''_{yy}(x,y)$

$$
\begin{aligned}
\frac{\partial E(w,b)}{\partial b} &= \frac{\partial}{\partial b}\left[\sum_{i=1}^{m}(y_i - wx_i - b)^2\right] \\
&= \sum_{i=1}^{m}\frac{\partial}{\partial b}(y_i - wx_i - b)^2 \\
&= \sum_{i=1}^{m} 2\cdot(y_i - wx_i - b)\cdot(-1) \\
&= 2\left(mb - \sum_{i=1}^{m}(y_i - wx_i)\right)
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
\frac{\partial^2 E_{(w,b)}}{\partial b^2} &= \frac{\partial}{\partial b}\left(\frac{\partial E_{(w,b)}}{\partial b}\right) \\
&= \frac{\partial}{\partial b}\left[2\left(mb - \sum_{i=1}^{m}(y_i - wx_i)\right)\right] \\
&= \frac{\partial}{\partial b}(2mb) \\
&= 2m
\end{aligned}
\tag{37}
$$

所　以，$C = 2m$

4. 求$AC - B^2$

$$
\begin{aligned}
AC - B^2 &= 4m\sum_{i=1}^{m} x_i^2 - 4m\cdot\bar{x}\cdot\sum_{i=1}^{m} x_i \\
&= 4m\left(\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i\bar{x}\right) \\
&= 4m\sum_{i=1}^{m}\left(x_i^2 - x_i\bar{x}\right)
\end{aligned}
\tag{38}
$$

又因为：

$$
\sum_{i=1}^{m} x_i\bar{x} = \bar{x}\sum_{i=1}^{m} x_i = \bar{x}\cdot m\cdot\frac{1}{m}\cdot\sum_{i=1}^{m} x_i = m\bar{x}^2 = \sum_{i=1}^{m}\bar{x}^2
\tag{39}
$$

所以：

$$m$$

$$
\begin{aligned}
AC-B^2 &= 4m\sum_{i=1}\left(x_i^2-x_i\bar{x}\right) \\
&= 4m\sum_{i=1}^{m}\left(x_i^2-x_i\bar{x}-x_i\bar{x}+x_i\bar{x}\right) \\
&= 4m\sum_{i=1}^{m}\left(x_i^2-x_i\bar{x}-x_i\bar{x}+\bar{x}^2\right) \\
&= 4m\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2 \\
&\geq 0
\end{aligned} \tag{40}
$$

因此，损失函数$\mathbf{E(w,b)}$是关于w和b的凸函数，问题得证。

1.3. 参数求解

1.3.1. 求偏置b

对损失函数$\mathbf{E(w,b)}$关于b求一阶偏导数：

$$
\frac{\partial E(w,b)}{\partial b}=2\left(mb-\sum_{i=1}^{m}\left(y_i-wx_i\right)\right) \tag{41}
$$

令一阶偏导数等于0解出b：

$$
\begin{gathered}
\frac{\partial E(w,b)}{\partial b}=2\left(mb-\sum_{i=1}^{m}\left(y_i-wx_i\right)\right)=0 \\
mb-\sum_{i=1}^{m}\left(y_i-wx_i\right)=0 \\
b=\frac{1}{m}\sum_{i=1}^{m}\left(y_i-wx_i\right) \\
b=\frac{1}{m}\sum_{i=1}^{m}y_i-w\cdot\frac{1}{m}\sum_{i=1}^{m}x_i=\bar{y}-w\bar{x}
\end{gathered} \tag{42}
$$

1.3.2. 求参数w

令一阶偏导数等于0解出w：

$$
\frac{\partial E(w,b)}{\partial w}=2\left(w\sum_{i=1}^{m}x_i^2-\sum_{i=1}^{m}\left(y_i-b\right)x_i\right)=0 \qquad w\sum_{i=1}^{m}x_i^2-\sum_{i=1}^{m}\left(y_i-b\right)x_i=0w\sum_{i=1}^{m}x_i^2=\sum_{i=1}^{m}y_ix_i-\sum_{i=1}^{m}bx_i \tag{43}
$$

将$b=\bar{y}-w\bar{x}$带入得：

$$
\begin{gathered}
w\sum_{i=1}^{m}x_i^2=\sum_{i=1}^{m}y_ix_i-\sum_{i=1}^{m}(\bar{y}-w\bar{x})x_i \\
w\sum_{i=1}^{m}x_i^2=\sum_{i=1}^{m}y_ix_i-\bar{y}\sum_{i=1}^{m}x_i+w\bar{x}\sum_{i=1}^{m}x_i \\
w\sum_{i=1}^{m}x_i^2-w\bar{x}\sum_{i=1}^{m}x_i=\sum_{i=1}^{m}y_ix_i-\bar{y}\sum_{i=1}^{m}x_i \\
w\left(\sum_{i=1}^{m}x_i^2-\bar{x}\sum_{i=1}^{m}x_i\right)=\sum_{i=1}^{m}y_ix_i-\bar{y}\sum_{i=1}^{m}x_i
\end{gathered} \tag{44}
$$

因为：

$$
\bar{y}\sum_{i=1}^{m}x_i=\frac{1}{m}\sum_{i=1}^{m}y_i\sum_{i=1}^{m}x_i=\bar{x}\sum_{i=1}^{m}y_i\bar{x}\sum_{i=1}^{m}x_i=\frac{1}{m}\sum_{i=1}^{m}x_i\sum_{i=1}^{m}x_i=\frac{1}{m}\left(\sum_{i=1}^{m}x_i\right)^2 \tag{45}
$$

所以：

$$
w=\frac{\sum_{i=1}^{m}y_ix_i-\bar{x}\sum_{i=1}^{m}y_i}{\sum_{i=1}^{m}x_i^2-\frac{1}{m}\left(\sum_{i=1}^{m}x_i\right)^2}=\frac{\sum_{i=1}^{m}y_i\left(x_i-\bar{x}\right)}{\sum_{i=1}^{m}x_i^2-\frac{1}{m}\left(\sum_{i=1}^{m}x_i\right)^2} \tag{46}
$$

记下来，对w进行向量化处理。

将$\frac{1}{m}\left(\sum_{i=1}^{m} x_i\right)^2 = \bar{x}\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} x_i\bar{x}$带入分母可得:

$$\begin{aligned} w= & \frac{\sum_{i=1}^{m} y_i\left(x_i-\bar{x}\right)}{\sum_{i=1}^{m} x_i^2-\sum_{i=1}^{m} x_i\bar{x}} \\ = & \frac{\sum_{i=1}^{m}\left(y_i x_i-y_i\bar{x}\right)}{\sum_{i=1}^{m}\left(x_{i=1}^2 x_i\bar{x}\right)} \end{aligned} \tag{47}$$

由于:

$$\text{由于}\begin{cases}\sum_{i=1}^{m} y_i\bar{x}=\bar{x}\sum_{i=1}^{m} y_i=\frac{1}{m}\sum_{i=1}^{m} x_i\sum_{i=1}^{m} y_i=\sum_{i=1}^{m} x_i\cdot\frac{1}{m}\cdot\sum_{i=1}^{m} y_i=\sum_{i=1}^{m} x_i\bar{y} \\ \sum_{i=1}^{m} y_i\bar{x}=\bar{x}\sum_{i=1}^{m} y_i=\bar{x}\cdot m\cdot\frac{1}{m}\cdot\sum_{i=1}^{m} y_i=m\bar{x}\bar{y}=\sum_{i=1}^{m}\bar{x}\bar{y} \\ \sum_{i=1}^{m} x_i\bar{x}=\bar{x}\sum_{i=1}^{m} x_i=\bar{x}\cdot m\cdot\frac{1}{m}\cdot\sum_{i=1}^{m} x_i=m\bar{x}^2=\sum_{i=1}^{m}\bar{x}^2\end{cases} \tag{48}$$

带入可得:

$$\begin{aligned} w= & \frac{\sum_{i=1}^{m}\left(y_i x_i-y_i\bar{x}\right)}{\sum_{i=1}^{m}\left(x_i^2-x_i\bar{x}\right)} \\ = & \frac{\sum_{i=1}^{m}\left(y_i x_i-y_i\bar{x}-y_i\bar{x}+y_i\bar{x}\right)}{\sum_{i=1}^{m}\left(x_i^2-x_i\bar{x}-x_i\bar{x}+x_i\bar{x}\right)} \\ = & \frac{\sum_{i=1}^{m}\left(y_i x_i-y_i\bar{x}-x_i\bar{y}+\bar{x}\bar{y}\right)}{\sum_{i=1}^{m}\left(x_i^2-x_i\bar{x}_i-x_i\bar{x}+\bar{x}^2\right)} \\ = & \frac{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2} \end{aligned} \tag{49}$$

定义向量:

$$\begin{aligned} \boldsymbol{x} &= \left(x_1, x_2, \cdots, x_m\right)^T \\ \boldsymbol{y} &= \left(y_1, y_2, \cdots, y_m\right)^T \\ \boldsymbol{x}_d &= \left(x_1-\bar{x}, x_2-\bar{x}, \ldots, x_m-\bar{x}\right)^T \\ \boldsymbol{y}_d &= \left(y_1-\bar{y}, y_2-\bar{y}, \ldots, y_m-\bar{y}\right)^T \end{aligned} \tag{50}$$

那么:

$$\begin{aligned} w &= \frac{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2} \\ &= \frac{\boldsymbol{x}_d^T\boldsymbol{y}_d}{\boldsymbol{x}_d^T\boldsymbol{x}_d} \end{aligned} \tag{51}$$

2. 多元线性回归

2.1. 模型定义

多元线性回归模型定义如下:

$$\begin{aligned} f(\boldsymbol{x_i}) &= w_1 x_{i1}+w_2 x_{i2}+\cdots+w_d x_{id}+b \\ &= w_1 x_{i1}+w_2 x_{i2}+\cdots+w_d x_{id}+w_{d+1}\cdot 1 \\ &= \left(w_1 \quad w_2 \quad \cdots \quad w_{d+1}\right)\cdot\begin{pmatrix} x_{i1} \\ x_{i2} \\ \cdots \\ x_{id} \\ 1 \end{pmatrix} \\ f(\hat{\boldsymbol{x}}_i) &= \hat{\boldsymbol{w}}^T\cdot\hat{\boldsymbol{x}}_i \\ & \text{其中，}\hat{\boldsymbol{w}}^T=\left(w_1 \quad w_2 \quad \cdots \quad w_{d+1}\right)\text{，}\hat{\boldsymbol{x}}_i^{\ T}=\left(x_{i1} \quad x_{i2} \quad \cdots \quad x_{id} \quad 1\right) \end{aligned} \tag{52}$$

令矩阵X为:

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1d} & 1 \\ x_{21} & x_{22} & \ldots & x_{2d} & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} & 1 \\ \boldsymbol{x}_2^{\mathrm{T}} & 1 \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{x}}_1^T \\ \hat{\boldsymbol{x}}_2^T \end{pmatrix} \tag{53}$$

$$
\begin{pmatrix} \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{md} & 1 \end{pmatrix}_{m\times d+1} \quad \begin{pmatrix} \vdots & \vdots \\ \boldsymbol{x}_m^{\mathrm{T}} & 1 \end{pmatrix}_{m\times d+1} \quad \begin{pmatrix} \vdots \\ \hat{\boldsymbol{x}}_m^{T} \end{pmatrix}_{m\times d+1}
$$

损失函数为：

$$
\begin{aligned}
E(\hat{\boldsymbol{w}}) &= \sum_{i=1}^{m}\left(y_i - f\left(\hat{\boldsymbol{x}}_i\right)\right)^2 \\
&= \sum_{i=1}^{m}\left(y_i - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_i\right)^2 \\
&= \begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_1 & y_2 - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_2 & \cdots & y_m - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_m \end{pmatrix} \cdot \begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_1 \\ y_2 - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_2 \\ \cdots \\ y_m - \hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_m \end{pmatrix} \\
&= (\boldsymbol{y} - \boldsymbol{X}\cdot\hat{\boldsymbol{w}})^T \cdot (\boldsymbol{y} - \boldsymbol{X}\cdot\hat{\boldsymbol{w}})
\end{aligned} \tag{54}
$$

2.2. 损失函数凹凸性判断

损失函数对变量求一阶偏导：

$$
\begin{aligned}
\frac{\partial E(\hat{\boldsymbol{w}})}{\partial\hat{\boldsymbol{w}}} &= \frac{\partial}{\partial\hat{\boldsymbol{w}}}\left[(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^T(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})\right] \\
&= \frac{\partial}{\partial\hat{\boldsymbol{w}}}\left[\left(\boldsymbol{y}^T - \hat{\boldsymbol{w}}^T\mathbf{X}^T\right)(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})\right] \\
&= \frac{\partial}{\partial\hat{\boldsymbol{w}}}\left[\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{y}^T\mathbf{X}\hat{\boldsymbol{w}} - \hat{\boldsymbol{w}}^T\mathbf{X}^T\boldsymbol{y} + \hat{\boldsymbol{w}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{w}}\right] \\
&= \frac{\partial}{\partial\hat{\boldsymbol{w}}}\left[-\boldsymbol{y}^T\mathbf{X}\hat{\boldsymbol{w}} - \hat{\boldsymbol{w}}^T\mathbf{X}^T\boldsymbol{y} + \hat{\boldsymbol{w}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{w}}\right] \\
&= -\frac{\partial\boldsymbol{y}^T\mathbf{X}\hat{\boldsymbol{w}}}{\partial\hat{\boldsymbol{w}}} - \frac{\partial\hat{\boldsymbol{w}}^T\mathbf{X}^T\boldsymbol{y}}{\partial\hat{\boldsymbol{w}}} + \frac{\partial\hat{\boldsymbol{w}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{w}}}{\partial\hat{\boldsymbol{w}}} \\
&= -\mathbf{X}^T\boldsymbol{y} - \mathbf{X}^T\boldsymbol{y} + \left(\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\mathbf{X}\right)\hat{w} \\
&= 2\mathbf{X}^T(\mathbf{X}\hat{w} - \boldsymbol{y})
\end{aligned} \tag{55}
$$

损失函数对变量求二阶偏导（Hessian矩阵）：

$$
\begin{aligned}
\frac{\partial^2 E(\hat{w})}{\partial\hat{w}\partial\hat{w}^T} &= \frac{\partial}{\partial\hat{w}}\left(\frac{\partial E_{\hat{w}}}{\partial\hat{w}}\right) \\
&= \frac{\partial}{\partial\hat{w}}\left[2\mathbf{X}^T(\mathbf{X}\hat{w} - \boldsymbol{y})\right] \\
&= \frac{\partial}{\partial\hat{w}}\left(2\mathbf{X}^T\mathbf{X}\hat{w} - 2\mathbf{X}^T\boldsymbol{y}\right) \\
&= 2\mathbf{X}^T\mathbf{X}
\end{aligned} \tag{56}
$$

说明：该Hessian矩阵无法保证为正定矩阵，但是，此处直接**假设该矩阵为正定矩阵**，否则无法继续推导。

2.3. 求$\hat{w}$

一阶偏导数为：

$$
\frac{\partial E(\hat{\boldsymbol{w}})}{\partial\hat{\boldsymbol{w}}} = 2\mathbf{X}^T(\mathbf{X}\hat{w} - \boldsymbol{y}) \tag{57}
$$

在Hessian矩阵为正定矩阵的假设下，令一阶偏导数为0，可得：

$$
\frac{\partial E(\hat{w})}{\partial\hat{w}} = 2\mathbf{X}^T(\mathbf{X}\hat{w} - \boldsymbol{y}) = 02\mathbf{X}^T\mathbf{X}\hat{w} - 2\mathbf{X}^T\boldsymbol{y} = 02\mathbf{X}^T\mathbf{X}\hat{w} = 2\mathbf{X}^T\boldsymbol{y}\hat{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y} \tag{58}
$$