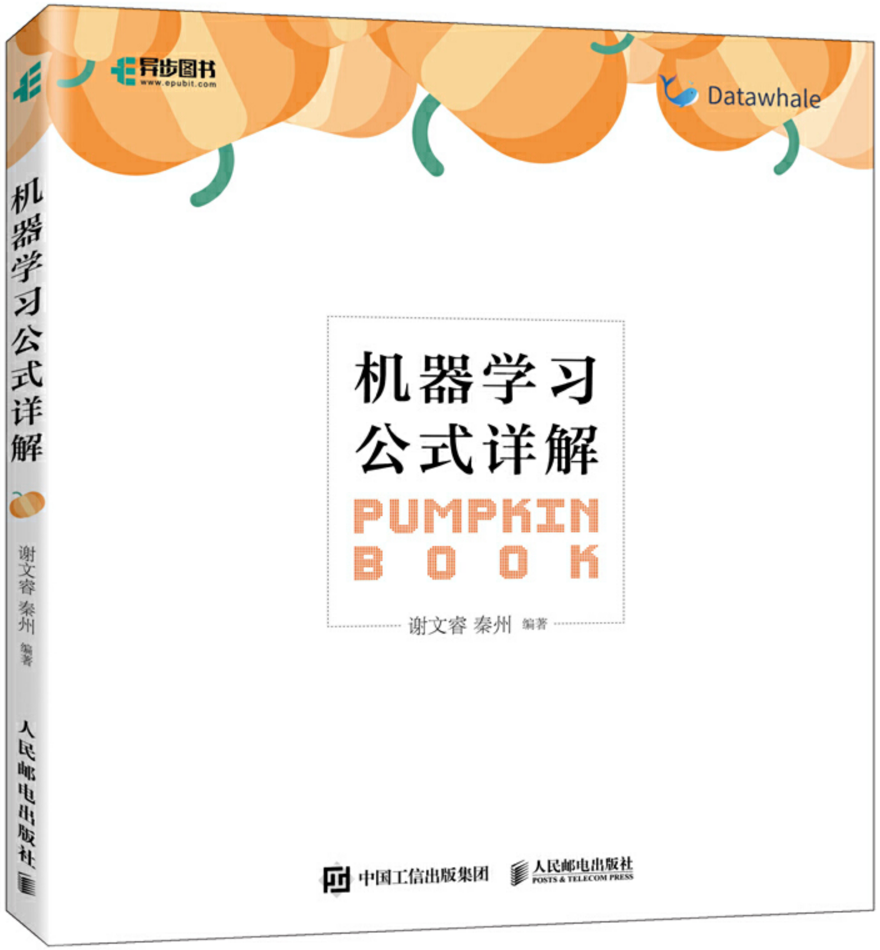
Datawhale 异步社区



《机器学习公式详解》（南瓜书）

第3章 多元线性回归

本节主讲：谢文睿

本节大纲

西瓜书对应章节：3.2

1. 由最小二乘法导出损失函数$E_{\hat{\boldsymbol{w}}}$
2. 求解$\hat{\boldsymbol{w}}$

导出$E_{\hat{\boldsymbol{w}}}$

将$\boldsymbol{w}$和b组合成$\hat{\boldsymbol{w}}$：

$$f(\boldsymbol{x}_i)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b$$

$$f(\boldsymbol{x}_i)=\begin{pmatrix} w_1 & w_2 & \cdots & w_d \end{pmatrix}\begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix}+b$$

$$f\left(\boldsymbol{x}_i\right)=w_1 x_{i1}+w_2 x_{i2}+\ldots+w_d x_{id}+b$$

$$f\left(\boldsymbol{x}_i\right)=w_1 x_{i1}+w_2 x_{i2}+\ldots+w_d x_{id}+w_{d+1}\cdot 1$$

导出$E_{\hat{\boldsymbol{w}}}$

$$f(\boldsymbol{x}_i)=\begin{pmatrix} w_1 & w_2 & \cdots & w_d & w_{d+1} \end{pmatrix}\begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \\ 1 \end{pmatrix}$$

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

导出$E_{\hat{\boldsymbol{w}}}$

Datawhale 异步社区

由最小二乘法可得

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

下面将其向量化以得到同西瓜书完全一致的形式

向量化$E_{\hat{\boldsymbol{w}}}$

$$E_{\hat{\boldsymbol{w}}}=\sum_{i=1}^{m}\left(y_i-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^2=\left(y_1-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1\right)^2+\left(y_2-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2\right)^2+\ldots+\left(y_m-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m\right)^2$$

$$E_{\hat{\boldsymbol{w}}}=\begin{pmatrix} y_1-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1 & y_2-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2 & \cdots & y_m-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m \end{pmatrix}\begin{pmatrix} y_1-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1 \\ y_2-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2 \\ \vdots \\ y_m-\hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m \end{pmatrix}$$

向量化$E_{\hat{\boldsymbol{w}}}$

其中

$$\begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1 \\ y_2 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2 \\ \vdots \\ y_m - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1 \\ \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2 \\ \vdots \\ \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} \hat{\boldsymbol{x}}_1^{\mathrm{T}}\hat{\boldsymbol{w}} \\ \hat{\boldsymbol{x}}_2^{\mathrm{T}}\hat{\boldsymbol{w}} \\ \vdots \\ \hat{\boldsymbol{x}}_m^{\mathrm{T}}\hat{\boldsymbol{w}} \end{pmatrix}$$

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \qquad \begin{pmatrix} \hat{\boldsymbol{x}}_1^{\mathrm{T}}\hat{\boldsymbol{w}} \\ \hat{\boldsymbol{x}}_2^{\mathrm{T}}\hat{\boldsymbol{w}} \\ \vdots \\ \hat{\boldsymbol{x}}_m^{\mathrm{T}}\hat{\boldsymbol{w}} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{x}}_1^{\mathrm{T}} \\ \hat{\boldsymbol{x}}_2^{\mathrm{T}} \\ \vdots \\ \hat{\boldsymbol{x}}_m^{\mathrm{T}} \end{pmatrix} \cdot \hat{\boldsymbol{w}} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} & 1 \\ \boldsymbol{x}_2^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_m^{\mathrm{T}} & 1 \end{pmatrix} \cdot \hat{\boldsymbol{w}} = \mathbf{X} \cdot \hat{\boldsymbol{w}}$$

向量化$E_{\hat{\boldsymbol{w}}}$

$$\begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1 \\ y_2 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2 \\ \vdots \\ y_m - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m \end{pmatrix} = \boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}}$$

$$E_{\hat{\boldsymbol{w}}} = \begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1 & y_2 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2 & \cdots & y_m - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m \end{pmatrix} \begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_1 \\ y_2 - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_2 \\ \vdots \\ y_m - \hat{\boldsymbol{w}}^{\mathrm{T}}\hat{\boldsymbol{x}}_m \end{pmatrix}$$

$$E_{\hat{\boldsymbol{w}}} = (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^{\mathrm{T}}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})$$

此即为公式3.9 $\underset{\hat{\boldsymbol{w}}}{\arg\min}$后面的部分

求解$\hat{\boldsymbol{w}}$

$$\hat{\boldsymbol{w}}^* = \underset{\hat{\boldsymbol{w}}}{\arg\min}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^{\mathrm{T}}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})$$

求解$\hat{\boldsymbol{w}}$仍然是一个多元函数求最值（点）的问题，同样也是凸函数求最值的问题。

推导思路：

1. 证明$E_{\hat{\boldsymbol{w}}} = (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^{\mathrm{T}}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})$是关于$\hat{\boldsymbol{w}}$的凸函数
2. 用凸函数求最值的思路求解出$\hat{\boldsymbol{w}}$

求解$\hat{\boldsymbol{w}}$

Datawhale 异步社区

求$E_{\hat{\boldsymbol{w}}}$的Hessian（海塞）矩阵$\nabla^2 E_{\hat{\boldsymbol{w}}}$，然后判断其正定性：

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

求解$\hat{\boldsymbol{w}}$

【标量-向量】的矩阵微分公式：设$\boldsymbol{x}\in\mathbb{R}^{n\times1}$, $f:\mathbb{R}^n\rightarrow\mathbb{R}$为关于$\boldsymbol{x}$的实值标量函数，则

$$\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}}=\begin{bmatrix}\frac{\partial f(\boldsymbol{x})}{\partial x_1}\\\frac{\partial f(\boldsymbol{x})}{\partial x_2}\\\vdots\\\frac{\partial f(\boldsymbol{x})}{\partial x_n}\end{bmatrix},\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}^{\mathrm{T}}}=\begin{pmatrix}\frac{\partial f(\boldsymbol{x})}{\partial x_1}&\frac{\partial f(\boldsymbol{x})}{\partial x_2}&\cdots&\frac{\partial f(\boldsymbol{x})}{\partial x_n}\end{pmatrix}$$

左侧为分母布局（默认），右侧为分子布局，仅差一个转置

推荐教材：张贤达.《矩阵分析与应用》

推荐手册：https://en.wikipedia.org/wiki/Matrix_calculus

求解$\hat{\boldsymbol{w}}$

$$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}}=-\frac{\partial \boldsymbol{y}^{\mathrm{T}} \mathbf{X} \hat{\boldsymbol{w}}}{\partial \hat{\boldsymbol{w}}}-\frac{\partial \hat{\boldsymbol{w}}^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \boldsymbol{y}}{\partial \hat{\boldsymbol{w}}}+\frac{\partial \hat{\boldsymbol{w}}^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbf{X} \hat{\boldsymbol{w}}}{\partial \hat{\boldsymbol{w}}}$$

根据矩阵微分公式$\frac{\partial \boldsymbol{x}^{\mathrm{T}} \boldsymbol{a}}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{a}, \frac{\partial \boldsymbol{x}^{\mathrm{T}} \mathbf{A} \boldsymbol{x}}{\partial \boldsymbol{x}}=\left(\mathbf{A}+\mathbf{A}^{\mathrm{T}}\right) \boldsymbol{x}$可得：

$$\begin{aligned}\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}} &=-\mathbf{X}^{T} \boldsymbol{y}-\mathbf{X}^{T} \boldsymbol{y}+\left(\mathbf{X}^{T} \mathbf{X}+\mathbf{X}^{T} \mathbf{X}\right) \hat{\boldsymbol{w}} \\ &=2 \mathbf{X}^{\mathrm{T}}(\mathbf{X} \hat{\boldsymbol{w}}-\boldsymbol{y})\text{公式3.10}\end{aligned}$$

求解$\hat{\boldsymbol{w}}$

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

根据矩阵微分公式$\dfrac{\partial \mathbf{A}\boldsymbol{x}}{\boldsymbol{x}} = \mathbf{A}^{\mathrm{T}}$可得：

$$
\nabla^2 E_{\hat{\boldsymbol{w}}} = 2\mathbf{X}^{\mathrm{T}}\mathbf{X}
$$

同西瓜书，假定$\mathbf{X}^{\mathrm{T}}\mathbf{X}$为正定矩阵，因此$E_{\hat{\boldsymbol{w}}}$是关于$\hat{\boldsymbol{w}}$的凸函数得证。

求解$\hat{\boldsymbol{w}}$

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

$$2\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{w}} - 2\mathbf{X}^{\mathrm{T}}\boldsymbol{y} = 0$$

$$2\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{w}} = 2\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$$

$$\hat{\boldsymbol{w}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$$

此即为公式3.11

预告

下一节：对数几率回归

西瓜书对应章节：3.3

结束语

欢迎加入【南瓜书读者交流群】，我们将在群里进行答疑、勘误、本次直播回放、本次直播PPT发放、下次直播通知等最新资源发放和活动通知。

加入步骤：

1. 关注公众号【Datawhale】，发送【南瓜书】三个字获取机器人“小豚”的微信二维码
2. 添加“小豚”为微信好友，然后对“小豚”发送【南瓜书】三个字即可自动邀请进群、