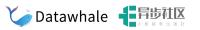


《机器学习公式详解》 (南瓜书)

第3章 多元线性回归

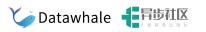
本节主讲: 谢文睿

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 $E_{\hat{m{w}}}$
- 2. 求解 $\hat{m{w}}$



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

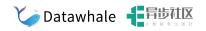
$$f(oldsymbol{x}_i) = oldsymbol{w}^{ ext{T}}oldsymbol{x}_i + b$$

$$f(oldsymbol{x}_i) = \left(egin{array}{cccc} w_1 & w_2 & \cdots & w_d \end{array}
ight) \left(egin{array}{c} x_{i1} \ x_{i2} \ dots \ x_{id} \end{array}
ight) + b$$

$$f\left(m{x}_i
ight) = w_1 x_{i1} + w_2 x_{i2} + ... + w_d x_{id} + b$$
 $f\left(m{x}_i
ight) = w_1 x_{i1} + w_2 x_{i2} + ... + w_d x_{id} + w_{d+1} \cdot 1$

$$f(oldsymbol{x}_i) = \left(egin{array}{cccc} w_1 & w_2 & \cdots & w_d & w_{d+1} \end{array}
ight) \left(egin{array}{c} x_{i1} \ x_{i2} \ dots \ x_{id} \ 1 \end{array}
ight)$$

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

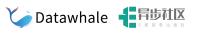


由最小二乘法可得

$$E_{\hat{oldsymbol{w}}} = \sum_{i=1}^m \left(y_i - f(\hat{oldsymbol{x}}_i)
ight)^2 = \sum_{i=1}^m \left(y_i - \hat{oldsymbol{w}}^{ ext{T}}\hat{oldsymbol{x}}_i
ight)^2.$$

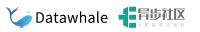
下面将其向量化以得到同西瓜书完全一致的形式

向量化 $E_{\hat{m{w}}}$



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

向量化 $E_{\hat{m{w}}}$

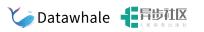


其中

$$egin{aligned} \left(egin{aligned} egin{aligned} y_1 - \hat{oldsymbol{w}}^{\mathrm{T}} \hat{oldsymbol{x}}_1 \ y_2 - \hat{oldsymbol{w}}^{\mathrm{T}} \hat{oldsymbol{x}}_2 \ dots \ y_m - \hat{oldsymbol{w}}^{\mathrm{T}} \hat{oldsymbol{x}}_m \end{aligned}
ight) = \left(egin{aligned} egin{aligned} egin{alig$$

$$oldsymbol{y} = \left(egin{array}{c} oldsymbol{y}_1 \ oldsymbol{y}_2 \ oldsymbol{arphi}_m \end{array}
ight), \qquad \left(egin{array}{c} \hat{oldsymbol{x}}_1^{
m T} \hat{oldsymbol{w}} \ oldsymbol{\hat{x}}_2^{
m T} \hat{oldsymbol{w}} \ oldsymbol{\hat{x}}_m^{
m T} \hat{oldsymbol{w}} \end{array}
ight) = \left(egin{array}{c} \hat{oldsymbol{x}}_1^{
m T} \ oldsymbol{\hat{x}}_2^{
m T} \ oldsymbol{\hat{x}}_m^{
m T} \end{array}
ight) \cdot \hat{oldsymbol{w}} = \left(egin{array}{c} oldsymbol{x}_1^{
m T} & 1 \ oldsymbol{\hat{x}}_2^{
m T} & 1 \ oldsymbol{\hat{x}}_m^{
m T} & 1 \end{array}
ight) \cdot \hat{oldsymbol{w}} = oldsymbol{X} \cdot \hat{oldsymbol{w}}$$

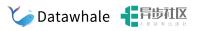
向量化 $E_{\hat{m{w}}}$



$$\left(egin{array}{c} oldsymbol{y}_1 - \hat{oldsymbol{w}}^{\mathrm{T}} \hat{oldsymbol{x}}_1 \ y_2 - \hat{oldsymbol{w}}^{\mathrm{T}} \hat{oldsymbol{x}}_2 \ dots \ y_m - \hat{oldsymbol{w}}^{\mathrm{T}} \hat{oldsymbol{x}}_m \end{array}
ight) = oldsymbol{y} - oldsymbol{X} \hat{oldsymbol{w}} \ y_m - oldsymbol{y}^{\mathrm{T}} \hat{oldsymbol{x}}_m \end{array}
ight)$$

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

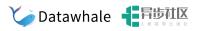
此即为公式3.9 arg min后面的部分



$$\hat{oldsymbol{w}}^* = rg\min_{\hat{oldsymbol{w}}} (oldsymbol{y} - \mathbf{X}\hat{oldsymbol{w}})^{\mathrm{T}} (oldsymbol{y} - \mathbf{X}\hat{oldsymbol{w}})^{\mathrm{T}}$$

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



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

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

求解 $\hat{m{w}}$



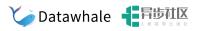
【标量-向量】的矩阵微分公式:设 $m{x}\in\mathbb{R}^{n\times 1}, f:\mathbb{R}^n\to\mathbb{R}$ 为关于 $m{x}$ 的实值标量函数,则

$$egin{aligned} rac{\partial f(oldsymbol{x})}{\partial oldsymbol{x}} = \left[egin{array}{c} rac{\partial f(oldsymbol{x})}{\partial x_1} \ drameth{arphi} \ drameth{arphi} \ rac{\partial f(oldsymbol{x})}{\partial x_2} \end{array}
ight], rac{\partial f(oldsymbol{x})}{\partial oldsymbol{x}^{
m T}} = \left(egin{array}{c} rac{\partial f(oldsymbol{x})}{\partial x_1} & rac{\partial f(oldsymbol{x})}{\partial x_2} & \cdots & rac{\partial f(oldsymbol{x})}{\partial x_n} \end{array}
ight) \end{aligned}$$

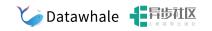
左侧为分母布局(默认),右侧为分子布局,仅差一个转置

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



$$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}} = -\frac{\partial \boldsymbol{y}^{\mathrm{T}} \mathbf{X} \hat{\boldsymbol{w}}}{\partial \hat{\boldsymbol{w}}} - \frac{\partial \hat{\boldsymbol{w}}^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \boldsymbol{y}}{\partial \hat{\boldsymbol{w}}} + \frac{\partial \hat{\boldsymbol{w}}^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbf{X} \hat{\boldsymbol{w}}}{\partial \hat{\boldsymbol{w}}}$$
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:
$$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}} = -\mathbf{X}^{T} \boldsymbol{y} - \mathbf{X}^{T} \boldsymbol{y} + (\mathbf{X}^{T} \mathbf{X} + \mathbf{X}^{T} \mathbf{X}) \hat{\boldsymbol{w}}$$
$$= 2\mathbf{X}^{\mathrm{T}} (\mathbf{X} \hat{\boldsymbol{w}} - \boldsymbol{y}) \triangle \vec{\mathbf{X}} 3.10$$

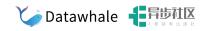


$$egin{aligned}
abla^2 E_{\hat{m{w}}} &= rac{\partial}{\partial \hat{m{w}}} \left(rac{\partial E_{\hat{m{w}}}}{\partial \hat{m{w}}}
ight) \ &= rac{\partial}{\partial \hat{m{w}}} \left[2 \mathbf{X}^{\mathrm{T}} (\mathbf{X} \hat{m{w}} - m{y})
ight] \ &= rac{\partial}{\partial \hat{m{w}}} \left(2 \mathbf{X}^{\mathrm{T}} \mathbf{X} \hat{m{w}} - 2 \mathbf{X}^{\mathrm{T}} m{y}
ight) \end{aligned}$$

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

$$abla^2 E_{\hat{m{w}}} = 2 \mathbf{X}^{ ext{T}} \mathbf{X}^{ ext{T}}$$

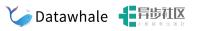
同西瓜书,假定 $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 为正定矩阵,因此 $E_{\hat{m{w}}}$ 是关于 $\hat{m{w}}$ 的凸函数得证。



$$egin{align} rac{\partial E_{\hat{oldsymbol{w}}}}{\partial \hat{oldsymbol{w}}} &= 2\mathbf{X}^{\mathrm{T}}(\mathbf{X}\hat{oldsymbol{w}} - oldsymbol{y}) = 0 \ & 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{oldsymbol{w}} - 2\mathbf{X}^{\mathrm{T}}oldsymbol{y} &= 0 \ & 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{oldsymbol{w}} = 2\mathbf{X}^{\mathrm{T}}oldsymbol{y} \ & \hat{oldsymbol{w}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}oldsymbol{y} \ \end{pmatrix}$$

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、

