deepshare.net
深度之眼

线性回归公式推导

Derivation of linear regression formula

本节大纲

Outline

先修内容：西瓜书3.1、3.2

1. 一元线性回归公式推导

2. 多元线性回归公式推导

多元线性回归公式推导

Derivation of simple linear regression formula

deepshare.net
深度之眼

求解权重 $\hat{w}$ 的公式推导

- 式3.9补充推导细节
- 式3.10补充推导细节
- 式3.11补充推导细节

求解权重$\hat{w}$的公式推导

Formula derivation for solving $\hat{w}$

deepshare.net
深度之眼

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 $\mathbf{0}$ 解出 $\hat{w}^*$

求解权重$\hat{\boldsymbol{w}}$的公式推导

deepshare.net
深度之眼

Formula derivation for solving $\hat{\boldsymbol{w}}$

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

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

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

$w_{d+1} \times 1$

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

求解权重$\hat{\boldsymbol{w}}$的公式推导

Formula derivation for solving $\hat{\boldsymbol{w}}$

$w^T x_i$

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

$\hat{w}^T$

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

$\hat{x}_i$

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

求解权重$\hat{\boldsymbol{w}}$的公式推导

Formula derivation for solving $\hat{\boldsymbol{w}}$

由最小二乘法导出损失函数 $E_{\hat{\boldsymbol{w}}}$：

$\hat{w}^T x_i$

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

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

求解权重 $\hat{\boldsymbol{w}}$ 的公式推导

deepshare.net
深度之眼

Formula derivation for solving $\hat{\boldsymbol{w}}$

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} & 1 \\ x_{21} & x_{22} & \dots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{md} & 1 \end{pmatrix}=\begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} & 1 \\ \boldsymbol{x}_2^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_m^{\mathrm{T}} & 1 \end{pmatrix}=\begin{pmatrix} \hat{\boldsymbol{x}}_1^T \\ \hat{\boldsymbol{x}}_2^T \\ \vdots \\ \hat{\boldsymbol{x}}_m^T \end{pmatrix}$$

$$\boldsymbol{y}=(y_1,y_2,\dots,y_m)^T \quad m\times 1$$

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

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

求解权重 $\hat{\boldsymbol{w}}$ 的公式推导

deepshare.net
深度之眼

Formula derivation for solving $\hat{\boldsymbol{w}}$

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

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

y $\quad d \times 1$

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

$X \cdot \hat{w} \qquad X \cdot \hat{w}$

又因为

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

$1 \times d+1$

$m \times d+1 \cdot d+1 \times 1 = m \times 1$

求解权重$\hat{\boldsymbol{w}}$的公式推导

Formula derivation for solving $\hat{\boldsymbol{w}}$

deepshare.net
深度之眼

所以

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

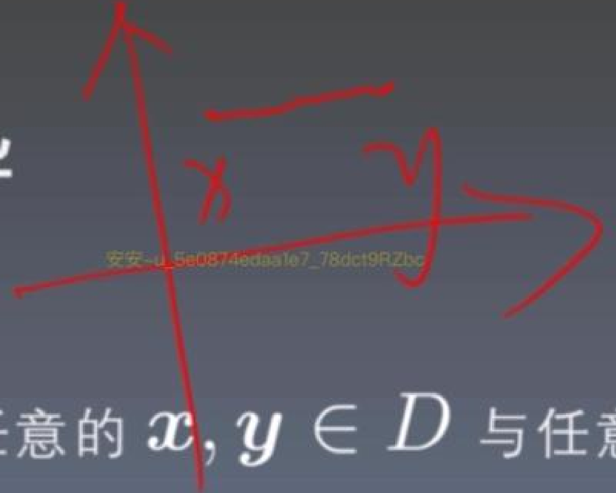
$$E_{\hat{\boldsymbol{w}}} = \begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_1 & y_2 - \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_2 & \cdots & y_d - \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_d \end{pmatrix} \begin{pmatrix} y_1 - \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_1 \\ y_2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

$$= (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^T (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})$$ 此即为式3.9 argmin后面的部分

求解权重 $\hat{\boldsymbol{w}}$ 的公式推导

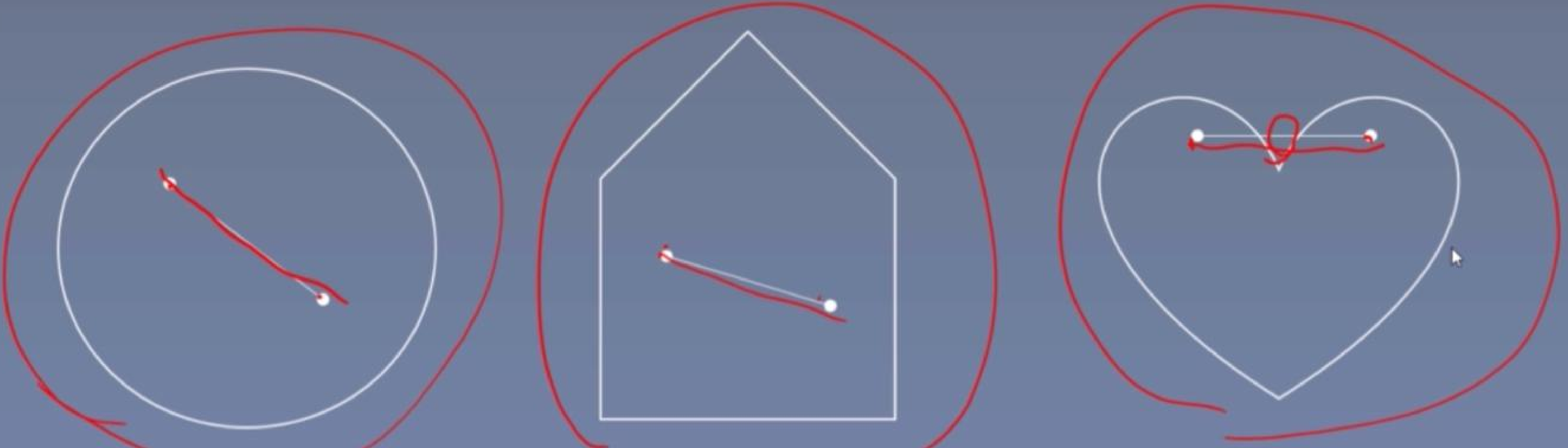
Formula derivation for solving $\hat{\boldsymbol{w}}$

deepshare.net
深度之眼



凸集定义：设集合 $D \in R^n$，如果对任意的 $\boldsymbol{x}, \boldsymbol{y} \in D$ 与任意的 $a \in [0,1]$,有 $a\boldsymbol{x} + (1-a)\boldsymbol{y} \in D$, 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。



凸集　　凸集　　非凸集

求解权重 $\hat{\boldsymbol{w}}$ 的公式推导

deepshare.net
深度之眼

Formula derivation for solving $\hat{\boldsymbol{w}}$

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

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

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

求解权重$\hat{\boldsymbol{w}}$的公式推导

Formula derivation for solving $\hat{\boldsymbol{w}}$

deepshare.net 深度之眼

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

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

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

求解权重$\hat{\boldsymbol{w}}$的公式推导

Formula derivation for solving $\hat{\boldsymbol{w}}$

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 $f(\boldsymbol{x})$ 在D上二阶连续可微,如果$f(\boldsymbol{x})$的Hessian矩阵$\nabla^2 f(\boldsymbol{x})$ 在D上是正定的,则 $f(\boldsymbol{x})$ 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 $f(\boldsymbol{x})$一阶连续可微,则 $\boldsymbol{x}^*$ 是全局解的充分必要条件是$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$,其中 $\nabla f(\boldsymbol{x})$为 $f(\boldsymbol{x})$ 关于$\boldsymbol{x}$的一阶导数(也称梯度)。

参考文献:王燕军,梁治安. 最优化基础理论与方法[M]. 复旦大学出版社, 2011.

求解权重$\hat{\boldsymbol{w}}$的公式推导

deepshare.net
深度之眼

Formula derivation for solving $\hat{\boldsymbol{w}}$

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

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

求解权重$\hat{w}$的公式推导

Formula derivation for solving $\hat{w}$

deepshare.net
深度之眼

【标量-向量】的矩阵微分公式为:

$$\frac{\partial y}{\partial \boldsymbol{x}} = \underbrace{\begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}}$$

(分母布局)[默认采用]

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{pmatrix}$$

(分子布局)

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

求解权重$\hat{w}$的公式推导

Formula derivation for solving $\hat{w}$

由【标量-向量】的矩阵微分公式可推得:

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

求解权重$\hat{w}$的公式推导

Formula derivation for solving $\hat{w}$

deepshare.net
深度之眼

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

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

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

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

$$= 2\mathbf{X}^T(\mathbf{X}\hat{w} - \boldsymbol{y})$$ 此即为式3.10

求解权重 $\hat{\boldsymbol{w}}$ 的公式推导

Formula derivation for solving $\hat{\boldsymbol{w}}$

deepshare.net
深度之眼

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

此即为Hessian矩阵

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

$$
\begin{aligned}
&(2X^TX)^T \\
&= 2(X^TX)^T \\
&= 2X^TX
\end{aligned}
$$

求解权重$\hat{w}$的公式推导

deepshare.net
深度之眼

Formula derivation for solving $\hat{w}$

对损失函数$E_{\hat{w}}$关于$\hat{w}$求一阶导数:

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

求解权重 $\hat{\boldsymbol{w}}$ 的公式推导

deepshare.net
深度之眼

Formula derivation for solving $\hat{\boldsymbol{w}}$

令一阶导数等于0解出 $\hat{\boldsymbol{w}}^*$:

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

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

$$(X^TX)^{-1} \quad \cancel{2}\mathbf{X}^T\mathbf{X}\hat{w} = \cancel{2}\mathbf{X}^T\boldsymbol{y} \qquad (X^TX)^{-1}X^Ty$$

$$(X^TX)^{-1}$$

$$E\ \hat{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$$ 此即为式3.11

结 语

在这次课程中，我们学习了西瓜书

多元线性回归的公式推导

那么在下次课程中，我们将会学习西瓜书

对数几率回归的公式推导