hw5_51215903008_陈诺

作业链接

https://www.wolai.com/9P1E94pcXNhhZfMXYoJCJm

习题部分

习题1.构建模型使得预测值与真实值的误差最小常用向量$2-$范数度量,
求解模型过程中需要计算梯度,求梯度 :

$$(1)f(\boldsymbol{A})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}-\boldsymbol{y}\|_2^2,求\frac{\partial f}{\partial \boldsymbol{A}}$$
$$(2)f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}-\boldsymbol{y}\|_2^2,术\frac{\partial f}{\partial \boldsymbol{x}}$$
$$其中\boldsymbol{A}\in\mathbb{R}^{m\times n},\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{b},\boldsymbol{y}\in\mathbb{R}^m$$

$$\begin{aligned}(1)f&=\frac{1}{2}(Ax+b-y)^{\top}(Ax+b-y)\\令z&=Ax+b-y\quad dz=dAx\\f&=\frac{1}{2}\operatorname{Tr}\left(z^{\top}z\right)\\df&=\frac{1}{2}\operatorname{Tr}\left(2z^{\top}dz\right)\\&=\operatorname{Tr}\left(z^{\top}dAx\right)\\&=\operatorname{Tr}\left(x(Ax+b-y)^{\top}dA\right)\end{aligned}$$
$$由df=Tr((\frac{\partial f}{\partial A})^{\top}dA),得\frac{\partial f}{\partial A}=(Ax+b-y)x^{\top}$$

$$\begin{aligned}(2)令z&=Ax+b-y\quad dz=Adx\\f&=\frac{1}{2}\operatorname{Tr}\left(z^{\top}z\right)\\df&=\frac{1}{2}\operatorname{Tr}\left(2z^{\top}dz\right)\\&=\operatorname{Tr}\left(z^{\top}Adx\right)\\&=\operatorname{Tr}\left((Ax+b-y)^{\top}Adx\right)\\\frac{\partial f}{\partial x}&=A^{\top}(Ax+b-y)\end{aligned}$$

$$\begin{aligned}(2')直接写\qquad f&=\frac{1}{2}\left(x^{\top}A^{\top}+(b-y)^{\top}\right)(Ax+b-y)\\\frac{\partial f}{\partial x}&=A^{\top}Ax+A^{\top}(b-y)\end{aligned}$$

习题2.求解$\frac{\partial \operatorname{Tr}\left(W^{-1}\right)}{\partial W}$,其中$W \in \mathbb{R}^{m \times m}$

$$
\begin{aligned}
\text{由} 0 = d\mathrm{I} &= d\left(WW^{-1}\right) \\
&= dW \cdot W^{-1} + W dW^{-1} \\
\therefore \quad dW^{-1} &= -W^{-1} dW W^{-1} \\
f &= \operatorname{Tr}\left(W^{-1}\right) \\
df &= \operatorname{Tr}\left(-W^{-1} dW W^{-1}\right) \\
&= \operatorname{Tr}\left(-W^{-1} W^{-1} dW\right) \\
&= \operatorname{Tr}\left(-W^{-2} dW\right) \\
\therefore \frac{\partial f}{\partial W} &= -\left(W^{-2}\right)^{\top}
\end{aligned}
$$

习题3.二次型是数据分析中常用函数,求$\frac{\partial x^T A x}{\partial x}, \frac{\partial x^T A x}{\partial A}$,其中$A \in \mathbb{R}^{m \times m}, x \in \mathbb{R}^m$

$$
\text{直接写} \quad \frac{\partial x^T A x}{\partial x} = (A + A^{\top} x), \frac{\partial x^T A x}{\partial A} = x x^{\top}
$$