Recitation 1_27

1 说明

学习本节内容前，最好复习一下国内的《线性代数》课程中有关矩阵运算的内容。

1.1 符号

公式中的符号通常代表的是一个矩阵，其中标量（普通数字）是1x1的矩阵，向量是mx1或1xn的矩阵。一般常用列向量代表一组数字，如x,y,λ一类的符号均代表列向量。

1.2 梯度

以如下函数为例

$$f(x)=wx=\begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} w_1x_1+w_2x_2 \\ w_3x_1+w_4x_2 \end{pmatrix}=\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

该函数对于矩阵x的梯度如下

$$\nabla_x f(x)=\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \end{pmatrix}=\begin{pmatrix} w_1 & w_3 \\ w_2 & w_4 \end{pmatrix}=w^T$$

即当$f(x)=wx$时，$\nabla_x f(x)=w^T$，维数改变时，结论仍然成立。

特别的，对于

$$f(x)=x^Tw=\begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}=\begin{pmatrix} w_1x_1+w_3x_2 \\ w_2x_1+w_4x_2 \end{pmatrix}$$

有

$$\nabla_x f(x)=\begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}=w$$

对于二次型

$$f(x)=\frac{1}{2}x^TQx \quad （Q\text{为实对称矩阵}）$$

有

$$\nabla_x f(x)=Qx$$

1.3 转置

特别注意的一点是，标量的转置等于它自身。当一个运算式的结果形状是1x1矩阵时，可以将整个运算式转置，这一操作不影响最终结果。

2 无约束条件的最小化问题

考虑如下的最小化问题

$$\min_x \quad \frac{1}{2}\|Ax-b\|_2^2 \ （\text{无约束条件}）$$

该最小化问题可经由以下步骤化简

$$\begin{aligned}\frac{1}{2}\|Ax-b\|_2^2 &= \frac{1}{2}(Ax-b)^T(Ax-b)\\ &= \frac{1}{2}(x^TA^T-b^T)(Ax-b)\\ &= \frac{1}{2}(x^TA^TAx-b^TAx-x^TA^Tb+b^Tb) \quad (b^Tb\text{与}x\text{无关，舍去})\\ &= \frac{1}{2}(x^TA^TAx-2b^TAx)\\ &= \frac{1}{2}x^TA^TAx-(A^Tb)^Tx\end{aligned}$$

可得如下的标准型

$$J(x)=\frac{1}{2}x^TA^TAx-(A^Tb)^Tx$$

其相对于x的梯度为

$$\nabla_x J(x)=A^TAx-A^Tb$$

令其等于0，可解得x的值为

$$x=(A^TA)^{-1}A^Tb$$

3 有约束条件的最小化问题

考虑如下最小化问题及约束条件

$$\begin{aligned}\min_x \quad & \frac{1}{2}\|x\|_2^2\\ s.t. \quad & Ax=b\end{aligned}$$

引入拉格朗日乘子

$$L(x,\lambda)=\frac{1}{2}x^Tx+\lambda^T(Ax-b)$$

为了解决最小化问题，我们需要解决如下的方程组

$$\nabla_x L(x,\lambda)=x+A^T\lambda=0 \quad (1)$$
$$Ax-b=0 \quad (2)$$

由(1)可得

$$x=-A^T\lambda$$

由(2)可得

$$\begin{aligned}Ax&=b\\ -AA^T\lambda&=b\\ \lambda&=-(AA^T)^{-1}b\end{aligned}$$

综上可得

$$x=A^T(AA^T)^{-1}b$$

4 通法

有如下最小化问题及约束条件

$$\begin{aligned}\min_{x} \quad & f(x) \\ s.t. \quad & C(x)=0\end{aligned}$$

引入拉格朗日乘子

$$L(x,\lambda)=f(x)+\lambda^T C(x)$$

得到如下的方程组

$$\nabla_x L(x,\lambda)=\nabla_x f(x)+\left(\frac{\partial C}{\partial x}\right)^T \lambda=0$$
$$C(x)=0$$

改写为如下形式

$$f(X)=f\begin{pmatrix} x \\ \lambda \end{pmatrix}=\begin{pmatrix} \nabla_x L(x,\lambda) \\ C(x) \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

则牛顿迭代法单步如下

$$X_{k+1}=X_k+\Delta X$$
$$\Delta X=\begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix}=-\begin{pmatrix} \nabla_x^2 L(x,\lambda) & \left(\frac{\partial C}{\partial x}\right)^T \\ \frac{\partial C}{\partial x} & 0 \end{pmatrix}\begin{pmatrix} \nabla_x L(x_k,\lambda_k) \\ C(x_k) \end{pmatrix}$$
$$\nabla_x^2 L(x,\lambda)=\nabla_x^2 f(x)+\frac{\partial}{\partial x}\left(\left(\frac{\partial C}{\partial x}\right)^T \lambda\right)$$