数据科学基础第一次作业

曾文正 U201715853 自动化校际 1701

请用课程中矩阵求导的方法求出下面优化问题的最优值：

$$\min_{x\in R^n} \|y-Ax\|_2^2+\lambda\|x\|_2^2$$

其中$y\in R^m$，$A\in R^{m\times n}$为给定向量、矩阵。

解：记$f(x)=\|y-Ax\|_2^2+\lambda\|x\|_2^2$，则有：

$$\begin{aligned}
f(x) &= \|y-Ax\|_2^2+\lambda\|x\|_2^2 \\
&= (y-Ax)^T(y-Ax)+\lambda x^T x \\
&= [y^T-(Ax)^T](y-Ax)+\lambda x^T x \\
&= (y^T-x^T A^T)(y-Ax)+\lambda x^T x \\
&= y^T y - y^T Ax - x^T A^T y + x^T A^T Ax + \lambda x^T x
\end{aligned}$$

接下来的推导中用到了老师上课讲的关于矩阵求导的三个引理:

$$\frac{\partial x^T Ax}{\partial x}=(A+A^T)x$$

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

$$\frac{\partial x^T x}{\partial x}=2x$$

其中，x、a 为列向量，A 为矩阵。

故有:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= 0-(y^T A)^T - A^T y + [A^T A + (A^T A)^T]x + 2\lambda x \\
&= 0 - A^T y - A^T y + (A^T A + A^T A)x + 2\lambda x \\
&= -2A^T y + 2A^T Ax + 2\lambda x \\
&= 2(A^T A+\lambda I)x - 2A^T y
\end{aligned}$$

对于最优值处，有

$$\frac{\partial f}{\partial x}=2(A^T A+\lambda I)x-2A^T y=0 \qquad ①$$

可得到极值点x_0:

$$2(A^T A+\lambda I)x=2A^T y$$

$$x_0=(A^T A+\lambda I)^{-1}A^T y$$

由①式，等式两边同时左乘x^T可得：

$$x_0{}^T(A^T A+\lambda I)x_0-x_0{}^T A^T y=0$$

故该优化问题的最优值为：

$$\begin{aligned}
f_{\min}(x)=f(x_0)&=\|y-Ax_0\|_2^2+\lambda\|x_0\|_2^2\\
&=(y-Ax_0)^T(y-Ax_0)+\lambda x_0{}^T\\
&=y^T y-y^T Ax_0-x_0{}^T A^T y+x_0{}^T A^T Ax_0+\lambda x_0{}^T x_0\\
&=y^T y-y^T Ax_0+[x_0{}^T(A^T A+\lambda I)x_0-x_0{}^T A^T y]\\
&=y^T y-y^T Ax_0+0\\
&=y^T y-y^T A(A^T A+\lambda I)^{-1}A^T y\\
&=y^T[\lambda I-A(A^T A+\lambda I)^{-1}A^T]y
\end{aligned}$$