常用向量与矩阵微分小结

完整版请见：

[机器学习中的矩阵、向量求导]

以及

[矩阵求导、几种重要的矩阵及常用的矩阵求导公式]

矩阵求导有两种布局方式，分子布局（Numerator layout）和分母布局（Denominator layout），两者求导的结果相差一个转置。一般来讲我们约定$x=(x_1,x_2,\ldots x_N)^T$，这是分母布局的方式。

以下未作特殊说明即为对变量$\mathbf{x}$求导。

标量对向量求导

- $\nabla\left(\boldsymbol{a}^T\boldsymbol{x}\right)=\nabla\left(\boldsymbol{x}^T\boldsymbol{a}\right)=\boldsymbol{a}$
- $\nabla\|\boldsymbol{x}\|_2^2=\nabla\left(\boldsymbol{x}^T\boldsymbol{x}\right)=2\boldsymbol{x}$
- $\nabla\left(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\right)=\left(A+A^T\right)\boldsymbol{x}$ (若$\boldsymbol{A}$为对称阵，则结果为$2\boldsymbol{A}\boldsymbol{x}$)
- $\nabla^2\left(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\right)=\left(A+A^T\right)$ (若$\boldsymbol{A}$为对称阵，则结果为$2\boldsymbol{A}$)
- $\nabla\left(\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{x}\right)=\boldsymbol{A}^T\boldsymbol{b}$

向量对向量求导

- $\nabla(\boldsymbol{x})=\boldsymbol{I}$
- $\nabla(\boldsymbol{A}\boldsymbol{x})=\boldsymbol{A}^{\boldsymbol{T}}$
- $\nabla\left(\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{A}\right)=\boldsymbol{A}$

以下未作特殊说明即为对变量X求导。

矩阵迹的求导

- 基本公式：$\nabla\operatorname{tr}\left(A^TX\right)=\nabla\operatorname{tr}\left(AX^T\right)=A$，$\nabla\operatorname{tr}(AX)=\nabla\operatorname{tr}(XA)=A^T$
- 核心公式：$\nabla\operatorname{tr}\left(XAX^TB\right)=B^TXA^T+BXA$
- $\nabla\boldsymbol{a}^TX\boldsymbol{b}=\boldsymbol{a}\boldsymbol{b}^T$
- $\nabla\boldsymbol{a}^TX^TX\boldsymbol{a}=2X\boldsymbol{a}\boldsymbol{a}^T$
- $\nabla(X\boldsymbol{a}-\boldsymbol{b})^T(X\boldsymbol{a}-\boldsymbol{b})=2(X\boldsymbol{a}-\boldsymbol{b})\boldsymbol{a}^T$
- $\nabla\|XA^T-B\|_F^2=2\left(XA^T-B\right)A$