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 $x=(x_1,x_2,\dots x_N)^T$,这是分母布局的方式。

以下未作特殊说明即为对变量x求导。

标量对向量求导

- $ullet \
 abla \left(oldsymbol{a}^Toldsymbol{x}
 ight) =
 abla \left(oldsymbol{x}^Toldsymbol{a}
 ight) = oldsymbol{a}$
- $\nabla (x^T A x) = (A + A^T) x$ (若A为对称阵,则结果为2Ax)
- $\nabla^2 (\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}) = (A + A^T)$ (若 \boldsymbol{A} 为对称阵,则结果为 $2\boldsymbol{A}$)
- ullet $\nabla \left(oldsymbol{b}^Toldsymbol{A}oldsymbol{x}
 ight) = oldsymbol{A}^Toldsymbol{b}$

向量对向量求导

- $\nabla (\boldsymbol{x}) = \boldsymbol{I}$
- ullet $\nabla \left(oldsymbol{A} oldsymbol{x}
 ight) = oldsymbol{A}^T$
- $ullet \
 abla \left(oldsymbol{x}^Toldsymbol{A}
 ight) = oldsymbol{A}$

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: $\nabla \operatorname{tr}(XAX^TB) = B^TXA^T + BXA$
- $\nabla \boldsymbol{a}^T X \boldsymbol{b} = \boldsymbol{a} \boldsymbol{b}^T$
- $\bullet \quad \nabla \boldsymbol{a}^T X^T X \boldsymbol{a} = 2 X \boldsymbol{a} \boldsymbol{a}^T$
- $\nabla (X\boldsymbol{a} \boldsymbol{b})^T (X\boldsymbol{a} \boldsymbol{b}) = 2(X\boldsymbol{a} \boldsymbol{b})\boldsymbol{a}^T$
- $\bullet \quad \nabla \|XA^T B\|_F^2 = 2\left(XA^T B\right)A$