Notes of Machine Learning

Kai Zhao

December 13, 2016

Contents

Part I

Supervised Learning

Chapter 1

Linear Regression

1.1 Matrix Derivatives

1.1.1 trace fact

1. $trAB = trBA$

proof:

Let A be a m-by-n matrix, let B be a n-by-m matrix.

$$trAB = \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}(\sum_{j=1}^{n} A_{ij}B_{ji}) \tag{1.1}$$

$$trBA = \sum_{i=1}^{n}(BA)_{ii} = \sum_{i=1}^{n}(\sum_{j=1}^{m} B_{ij}A_{ji}) \tag{1.2}$$

$$\begin{aligned}
trBA &= \sum_{i=1}^{n}(\sum_{j=1}^{m} B_{ij}A_{ji}) \\
&= \sum_{j=1}^{m}(\sum_{i=1}^{n} B_{ij}A_{ji}) \\
&= \sum_{j=1}^{m}(\sum_{i=1}^{n} A_{ji}B_{ij}) \\
&= \sum_{i=1}^{m}(\sum_{j=1}^{n} A_{ij}B_{ji}) \\
&= trAB
\end{aligned} \tag{1.3}$$

2. $\nabla_A trAB = B^T$

proof:

Let A be a m-by-n matrix, let B be a n-by-m matrix.

$$
\begin{aligned}
\nabla_A trAB &= \begin{bmatrix} \frac{trAB}{\partial A_{11}} \cdots \frac{trAB}{\partial A_{1n}} \\ \cdots \\ \cdots \\ \cdots \\ \frac{trAB}{\partial A_{m1}} \cdots \frac{trAB}{\partial A_{mn}} \end{bmatrix} \\
&= \begin{bmatrix} B_{11} \ldots B_{n1} \\ \ldots \\ \ldots \\ \ldots \\ B_{1m} \ldots B_{nm} \end{bmatrix} \\
&= B^T
\end{aligned} \tag{1.4}
$$