Notes of Machine Learning

Kai Zhao

January 18, 2017

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.

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

3. $\nabla_{A^T} f(A) = (\nabla_A f(A))^T$

proof:

Let A be a m-by-n matrix.

$$
\nabla_{A^T} f(A) = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} \cdots \frac{\partial f(A)}{\partial A_{m1}} \\ \cdots \\ \cdots \\ \cdots \\ \frac{\partial f(A)}{\partial A_{1n}} \cdots \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix} \tag{1.5}
$$

$$
\begin{aligned}
(\nabla_A f(A))^T &= \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} \cdots \frac{\partial f(A)}{\partial A_{1n}} \\ \cdots \\ \cdots \\ \cdots \\ \frac{\partial f(A)}{\partial A_{m1}} \cdots \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}^T \\
&= \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} \cdots \frac{\partial f(A)}{\partial A_{m1}} \\ \cdots \\ \cdots \\ \cdots \\ \frac{\partial f(A)}{\partial A_{1n}} \cdots \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix} \\
&= \nabla_{A^T} f(A)
\end{aligned} \tag{1.6}
$$

Chapter 2

Boosting

2.1 Boosting

2.1.1 Reference

http://www.cnblogs.com/wentingtu/archive/2011/12/15/2289550.html

Wikipedia: Boosting

2.1.2 Definition

Boosting is a family of machine learning algorithms which convert **weak learners** to **strong ones**.

2.2 Gradient boosting

2.2.1 Reference

http://www.cnblogs.com/wentingtu/archive/2011/12/15/2289550.html

2.2.2 Definition

Gradient boosting is a method of boosting. It is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.

It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.