

ML & LinAlg Math Cheat Sheet

October 24, 2017

Contents

1	Notation	1
2	Derivative	1
2.a	Vector Gradient	1
3	Determinant Operator	2
3.a	Random Properties	2
4	Trace Operator	2
4.a	Derivatives	2
4.a.i	$\nabla_{\mathbf{x}} \text{tr}(\mathbf{x}\mathbf{x}^\top \mathbf{A}) = 2\mathbf{x}^\top \mathbf{A}$	2
4.b	Relation to Determinant	2

1 Notation

Vectors are column vectors denoted by lower-case bolded variables, such that

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}.$$

A row vector is denoted $\mathbf{x}^\top = [x_1 \dots x_N]$. A matrix is indicated by a bolded upper-case variable, such that an $N \times M$ matrix is

$$\mathbf{A} = \{a_{ij}\} = [\mathbf{a}_1 \dots \mathbf{a}_M] = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_N^\top \end{bmatrix} = \begin{bmatrix} a_{1,1} & \dots & a_{1,M} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \dots & a_{N,M} \end{bmatrix}.$$

2 Derivative

2.a Vector Gradient

$$\nabla_{\mathbf{x}} \mathbf{y} = \left[\frac{\partial \mathbf{y}}{\partial x_1}, \dots, \frac{\partial \mathbf{y}}{\partial x_N} \right] \quad (1)$$

3 Determinant Operator

3.a Random Properties

For scalar c and $N \times N$ identity matrix I ,

$$\det(cI) = c^N.$$

4 Trace Operator

4.a Derivatives

4.a.i $\nabla_{\mathbf{x}} \text{tr}(\mathbf{x}\mathbf{x}^\top \mathbf{A}) = 2\mathbf{x}^\top \mathbf{A}$

For square $N \times N$ matrix \mathbf{A} ,

$$\nabla_{\mathbf{x}} \text{tr}(\mathbf{x}\mathbf{x}^\top \mathbf{A}) = \frac{d}{d\mathbf{x}} \text{tr}(\mathbf{x}\mathbf{x}^\top \mathbf{A}) = \frac{d}{d\mathbf{x}} \sum_i^N \sum_k^N x_i x_k a_{ik}.$$

Recall eq. (1), and consider for any $j \in \{1, \dots, N\}$:

$$\begin{aligned} \frac{\partial}{\partial x_j} \sum_i^N \sum_k^N x_i x_k a_{ik} &= x_1 a_{j,1} + x_2 a_{j,2} + \dots + \frac{\partial}{\partial x_j} \sum_k^N x_j x_k a_{jk} + \dots + x_N a_{j,N} \\ &= \sum_i^N x_i a_{ji} - x_j a_{jj} + \sum_k^N x_k a_{jk} - x_j a_{jj} + \frac{\partial}{\partial x_j} x_j x_j a_{jj} \\ &= 2 \sum_i^N x_i a_{ji} - 2x_j a_{jj} + 2x_j a_{jj} \\ &= 2\mathbf{x}^\top \mathbf{a}_j. \end{aligned}$$

This equally applies for any j in $1 \dots N$, and so for the full gradient:

$$\nabla_{\mathbf{x}} \text{tr}(\mathbf{x}\mathbf{x}^\top \mathbf{A}) = \frac{d}{d\mathbf{x}} \sum_i^N \sum_k^N x_i x_k a_{ik} = [2\mathbf{x}^\top \mathbf{a}_1 \dots 2\mathbf{x}^\top \mathbf{a}_N] = 2\mathbf{x}^\top \mathbf{A}. \quad (2)$$

4.b Relation to Determinant