

Properties of the Trace

Marc Deisenroth

March 8, 2022

The trace $\text{tr}(\mathbf{A})$ of a square matrix $\mathbf{A} \in \mathbb{R}^{D \times D}$ is defined as the sum of the diagonal elements, i.e.,

$$\text{tr}(\mathbf{A}) = \sum_{d=1}^D A_{dd}, \quad \mathbf{A} \in \mathbb{R}^{D \times D}. \quad (1)$$

It has the following properties:

$$\text{tr}(\alpha) = \alpha, \quad \alpha \in \mathbb{R} \quad (2)$$

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}), \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{D \times D} \quad (3)$$

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB}), \quad \mathbf{A} \in \mathbb{R}^{D \times E}, \mathbf{B} \in \mathbb{R}^{E \times F}, \mathbf{C} \in \mathbb{R}^{F \times D} \quad (4)$$

Exploiting these properties (jumping toward the end of the PCA loss function reformulation [1]), we get

$$\sum_{j=M+1}^D \underbrace{\mathbf{b}_j^\top \mathbf{S} \mathbf{b}_j}_{\in \mathbb{R}} \stackrel{(2)}{=} \sum_{j=M+1}^D \text{tr}(\mathbf{b}_j^\top \mathbf{S} \mathbf{b}_j) \quad (5)$$

$$\stackrel{(4)}{=} \sum_{j=M+1}^D \text{tr}(\mathbf{S} \mathbf{b}_j \mathbf{b}_j^\top) \quad (6)$$

$$\stackrel{(4)}{=} \sum_{j=M+1}^D \text{tr}(\mathbf{b}_j \mathbf{b}_j^\top \mathbf{S}) \quad (7)$$

$$\stackrel{(3)}{=} \text{tr} \left(\sum_{j=M+1}^D \mathbf{b}_j \mathbf{b}_j^\top \mathbf{S} \right) \quad (8)$$

$$= \text{tr} \left(\left(\sum_{j=M+1}^D \mathbf{b}_j \mathbf{b}_j^\top \right) \mathbf{S} \right) \quad (9)$$

where in the last step we used the fact that \mathbf{S} is independent of j , so that we could move it out of the sum. For further details on the trace and the reformulation of the loss function in PCA, we refer to [1], Chapter 10.3.

References

- [1] M. P. Deisenroth, A. A. Faisal, and C. S. Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020.