Machine Learning

Appendix

Yannick Le Cacheux

CentraleSupélec - Université Paris Saclay

September 2024

Matrix calculus

We can define:

- Derivatives of scalars with respect to vectors (*i.e.* gradients):

$$\text{For } a \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^N, \quad \frac{\partial a}{\partial \mathbf{x}} \in \mathbb{R}^N \quad \text{and} \quad \boxed{\left(\frac{\partial a}{\partial \mathbf{x}}\right)_i = \frac{\partial a}{\partial x_i}}$$

- But also derivatives of vectors with respect to scalars:

$$\text{For } \mathbf{a} \in \mathbb{R}^N, x \in \mathbb{R}, \quad \frac{\partial \mathbf{a}}{\partial x} \in \mathbb{R}^N \quad \text{and} \quad \boxed{\left(\frac{\partial \mathbf{a}}{\partial x}\right)_i = \frac{\partial a_i}{\partial x}}$$

- Or derivatives of vectors w.r.t. vectors:

$$\text{For } \mathbf{a} \in \mathbb{R}^M, \mathbf{b} \in \mathbb{R}^N, \quad \frac{\partial \mathbf{a}}{\partial \mathbf{b}} \in \mathbb{R}^{M \times N} \quad \text{and} \quad \boxed{\left(\frac{\partial \mathbf{a}}{\partial \mathbf{b}}\right)_{ij} = \frac{\partial a_i}{\partial b_j}}$$

- etc

Matrix calculus

We can then prove:

- If $\mathbf{a}$ is constant with respect to $\mathbf{x}$:

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top \mathbf{a}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{a}^\top \mathbf{x}) = \mathbf{a} \tag{1}$$

- For matrices $\mathbf{A}(x)$ and $\mathbf{B}(x)$ that depend on x:

$$\frac{\partial}{\partial x}(\mathbf{AB}) = \frac{\partial \mathbf{A}}{\partial x}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial x} \tag{2}$$

Exercise[1]: prove that:

$$\frac{\partial \mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\mathbf{A}^{-1}$$

[1] In homework assignment 1. Hint: use the fact that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and equation (2)

Matrix calculus

- For any 3 matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ that do not depend on a 4$^{\text{th}}$ matrix $\mathbf{X}$, and defined such that the equation below makes sense, we can prove:

Lemma

$$\frac{\partial \|\mathbf{AWB} + \mathbf{C}\|_2^2}{\partial \mathbf{X}} = 2\mathbf{A}^\top(\mathbf{AXB} + \mathbf{C})\mathbf{B}^\top$$

- This is a generic result whose special cases we will often be useful in this course.