Machine Learning

Appendix

Yannick Le Cacheux

CentraleSupélec - Université Paris Saclay

September 2024

Table of Contents

1. Notations
2. Probability reminders
3. Linear algebra reminders
4. Matrix calculus

Outline

1. Notations
2. Probability reminders
3. Linear algebra reminders
4. Matrix calculus

Notations

In general, in this class:

- A lowercase, non bold letter is a scalar:

$$x \in \mathbb{R}$$

- A lowercase, bold letter is a vector:

$$\mathbf{x} \in \mathbb{R}^N$$

- An uppercase, bold letter is a matrix:

$$\mathbf{X} \in \mathbb{R}^{M \times N}$$

Outline

1. Notations
2. Probability reminders
3. Linear algebra reminders
4. Matrix calculus

Expectation and (co)variance

$$\text{cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if and only if X and Y are independent, *i.e.*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \iff X \perp\!\!\!\perp Y$$

If $X = Y$ then

$$\text{cov}[X,X] = \text{var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Outline

1 Notations

2 Probability reminders

3 Linear algebra reminders

4 Matrix calculus

Linear algebra

The projection $\text{Proj}_{\mathbf{b}}(\mathbf{a})$ of vector $\mathbf{a}$ onto vector $\mathbf{b}$:

- Has amplitude
$$||\mathbf{a}||\cos(\theta)$$
where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$
- Has direction
$$\frac{\mathbf{b}}{||\mathbf{b}||}$$
(unit vector in the direction of $\mathbf{b}$)
- The dot product between $\mathbf{a}$ and $\mathbf{b}$ is
$$\mathbf{a}^\top \mathbf{b} = ||\mathbf{a}||||\mathbf{b}||\cos(\theta)$$
- Consequently,
$$\text{Proj}_{\mathbf{b}}(\mathbf{a}) = \frac{\mathbf{a}^\top \mathbf{b}}{||\mathbf{b}||^2} \cdot \mathbf{b}$$

Outline

1 Notations

2 Probability reminders

3 Linear algebra reminders

4 Matrix calculus

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 $\mathbf{x}$ ($\mathbf{a} \neq \mathbf{a}(\mathbf{x})$):

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

- 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 for $\mathbf{A}(x)$:

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

[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.

Outline

1 Notations

2 Probability reminders

3 Linear algebra reminders

4 Matrix calculus

Constrained optimization

Coming soon