Université d'Aix-Marseille
Master Intelligence Artificielle
Année universitaire 2025–2026

Mathématiques pour l'Intelligence Artificielle

Notes de cours – CM6

Rédigé par :
GUELLIL Rayane
BRAHIM Haroun Hassan
ANLAOUDINE Moindze Boina

Marseille, le 30 septembre 2025

Table des matières

1 Structure typique d'un problème d'apprentissage automatique

Un problème d'apprentissage automatique comprend généralement :

— des **données d'entraînement** et une **fonction de perte**,

— un **espace d'hypothèses** décrivant les modèles possibles,

— une **procédure d'entraînement** cherchant à minimiser la perte.

L'objectif est d'obtenir une performance suffisante pour l'application visée, puis d'évaluer le modèle dans son environnement de production.

2 Rappels de probabilités et espérance

Espérance (cas discret et continu)

$$\text{Discret} : E[X] = \sum_x x\, p(x)$$
$$\text{Continu} : E[X] = \int x\, p(x)\, dx$$

Pour une fonction $f(X)$:

$$E[f(X)] = \begin{cases} \sum_x f(x)p(x) & \text{(discret)} \\ \int f(x)p(x)\, dx & \text{(continu)} \end{cases}$$

Cas multidimensionnel

Si $X = (X_1, \ldots, X_n)$:

$$E[f(X)] = \int \cdots \int f(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

Exemple – Loi de l'espérance totale

Un téléphone tient en moyenne 12h avec la batterie A et 8h avec la batterie B. 80% des téléphones ont la batterie A. On a alors :

$$E[T] = E[T|A]P(A) + E[T|B]P(B) = 12 \times 0.8 + 8 \times 0.2 = 11.2\, h$$

Ainsi, l'espérance de la durée de vie moyenne est de **11,2 heures**.

3 Moments et covariances

— Moment d'ordre k : $E[X^k]$

— Moment centré d'ordre k : $E[(X - E[X])^k]$

— Variance : $Var[X] = E[(X - E[X])^2]$

— Covariance : $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$

Propriétés

$$
\begin{aligned}
E\left[\sum_i a_i X_i\right] &= \sum_i a_i E[X_i] \quad \text{(linéarité)} \\
Cov\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) &= \sum_i \sum_j a_i b_j \, Cov(X_i, Y_j) \quad \text{(bilinéarité)}
\end{aligned}
$$

Exemple – Variance d'une somme

Pour $S = X + Y$, on a :

$$Var[S] = Var[X] + Var[Y] + 2\, Cov(X, Y)$$

Si X et Y sont indépendantes, $Cov(X, Y) = 0$, donc $Var[S] = Var[X] + Var[Y]$.

4 Espérance et variance conditionnelles

Espérance conditionnelle

$$
\begin{aligned}
\text{Discret : } E[X|Y = y] &= \sum_x x\, p(x|y) \\
\text{Continu : } E[X|Y = y] &= \int x\, p(x|y)\, dx
\end{aligned}
$$

Lois totales

$$
\begin{aligned}
E[X] &= E[E[X|Y]] \\
Cov[X, Y] &= Cov[E[X|Z], E[Y|Z]] + E[Cov[X, Y|Z]]
\end{aligned}
$$

Variance conditionnelle

$$Var[X|Y] = Cov(X, X|Y)$$

5 Cadre général de l'analyse statistique

On cherche à estimer une fonctionnelle $f(P)$ à partir d'observations $O \sim P$.
Un **estimateur** $\hat{f}(O)$ vise à approximer $f(P)$.

Éléments du cadre

- Observations : $O \sim P$
- Distribution : $P \in \mathcal{P}$
- Estimateur : $\hat{f} : O \to E$
- Fonctionnelle : $f : \mathcal{P} \to E$

Exemple – Moyenne empirique

Soient $X_1, \ldots, X_n$ i.i.d. de moyenne μ et variance σ^2. L'estimateur de la moyenne empirique est :

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Alors :

$$E[\hat{\mu}] = \mu \quad \text{et} \quad Var[\hat{\mu}] = \frac{\sigma^2}{n}$$

$\Rightarrow$ $\hat{\mu}$ est un **estimateur sans biais** et sa variance diminue quand n augmente.

6 Qualité statistique d'un estimateur ponctuel

Trois critères

$$\begin{aligned} \text{Biais :} \quad & b_P(\hat{f}) = E_{O\sim P}[\hat{f}(O)] - f(P) \\ \text{Variance :} \quad & Var_P(\hat{f}) = Var_{O\sim P}[\hat{f}(O)] \\ \text{Risque :} \quad & R_P(\hat{f}) = E_{O\sim P}[\ell(f(P), \hat{f}(O))] \end{aligned}$$

Exemple – Biais d'un estimateur de variance

Pour un échantillon i.i.d. $X_1, \ldots, X_n$:

$$\hat{V} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

est un **estimateur biaisé** de la variance réelle σ^2. En effet :

$$E[\hat{V}] = \frac{n-1}{n} \sigma^2$$

On corrige ce biais en utilisant :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

$\Rightarrow$ S^2 est un **estimateur non biaisé** de σ^2.

Interprétation

— Le **biais** mesure l'écart moyen entre la prédiction et la vraie valeur.

— La **variance** évalue la sensibilité de l'estimateur aux données d'entraînement.

— Le **risque** combine biais et variance selon une fonction de perte ℓ.

Un bon estimateur minimise le **compromis biais–variance**.