
Machine learning cheatsheet

Alexis GRACIAS

13 janvier 2026

Table des matières

1	Formulation mathématique d'un <i>dataset</i>	2
2	Métriques	3
2.1	Régression	3
2.2	Classification	3
3	Synthèse des modèles et leurs usages	5

1 | Formulation mathématique d'un *data-set*

2 | Métriques

2.1 Régression

- **Mean Absolute Error :**

$$MSE(Y) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.1)$$

- **Root Mean Squarred Error :**

$$MSE(Y) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.2)$$

•

2.2 Classification

- **Accuracy score :**

$$\text{Accuracy} = \frac{VP + VN}{VP + VN + FN + FP} = 1 - \text{exactitude} \quad (2.3)$$

Avec :

— VP : Vrai négatif

— FP : Faux positifs

— FN : Faux négatifs

— VP : Vrais positifs

- **Matrice de confusion**

Une matrice de confusion recense le nombre de faux positifs, faux négatifs, vrai positifs et vrai négatifs sous forme de tableau. Les éléments de cette matrice 2×2 veulent dire ceci :

- Vrai négatif (Réal 0, Prédit 0)
- Faux positifs (Réal 0, Prédit 1)
- Faux négatifs (Réal 1, Prédit 0)
- Vrais positifs (Réal 1, Prédit 1)

La prédiction est la valeur que le modèle prédit tandis que réel est la valeur réelle, qui peut être observée à **posteriori** Par exemple, une matrice de confusion peut ressembler à ça :

	Prédit : 0	Prédit : 1
Réal : 0	87	13
Réal : 1	18	61

TABLE 2.1 – Exemple de matrice de confusion

Ici, le modèle a prédit que $87 + 18 = 105$ personnes n'ont pas survécues à l'accident sur les $87 + 13 = 100$ personnes qui n'ont réellement pas survécues. De même, le modèle prédit que $13 + 61 = 74$ personnes on survécues sur $10 + 61 = 71$ qui ont réellement survécues.

3 | Synthèse des modèles et leurs usages

Modèle	Quand l'utiliser	Paradigme	Formule	Quantité à minimiser	Préparation des données
Régression linéaire	Relation linéaire entre X_i et X_j	Régression statistique	$\hat{y} = X\beta + \varepsilon$	$\ y - X\beta\ ^2$	Standardisation
Régression Ridge	Régression avec pénalisation L_2 (sélection de variables)	Régression statistique	$\hat{y} = X\beta + \varepsilon$	$\ y - X\beta\ ^2 + \lambda\ \beta\ ^2$	Standardisation
Régression Lasso	Régression avec pénalisation L_1 (sélection de variables avec annulation)	Régression statistique	$\hat{y} = X\beta + \varepsilon$	$\ y - X\beta\ ^2 + \lambda\ \beta\ _1$	Standardisation
Elastic Net	Régression Ridge + Lasso	Régression statistique	$\hat{y} = X\beta + \varepsilon$	$\ y - X\beta\ ^2 + \lambda(\alpha\ \beta\ _1 + (1 - \alpha)\ \beta\ _2^2)$	Standardisation
Régression logistique	Classification binaire	Classification binaire	$\hat{y} = \sigma(X\beta)$	$-\sum y \log(\hat{y})$	Standardisation
Régression logistique multinomiale	Classification	Classification			Standardisation
Régression polynomiale	Régression non linéaire	Régression statistique	$\hat{y} = \sum_{k=0}^d \beta_k x^k$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	Standardisation
Régression splines	Régression non linéaire, robuste, locale	Régression statistique			-
Régression additive généralisée (GAM)	Régression fortement non linéaire, robuste, locale	Régression statistique			-
Régression isotonique	Régression non linéaire monotone	Régression statistique			Standardisation
Régression Généralisée (GLM)	Régression non linéaire	Régression statistique			Standardisation
Quantile regression	Régression fortement non linéaire, robuste, présence d'hétéroscédasticité	Régression statistique		pinball loss	Standardisation
Processus gaussiens (GP)	Régression fortement non linéaire	Régression statistique			-
Huber regression	Régression fortement non linéaire, robuste, insensible aux outliers	Régression statistique		huber loss	-
Régression PLS	Régression non linéaire avec analyse des composantes principales	Régression non supervisée			Standardisation
KNN regression	Régression locale / ponctuelle avec beaucoup de données	Régression statistique	$\hat{y} = \frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} y_i$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	Standardisation

TABLE 3.1 – Synthèse des modèles de Machine Learning - régression

Modèle	Quand l'utiliser	Paradigme	Formule	Quantité à minimiser	Préparation des données
SVM (Support Vector Machine)	Classification fortement non linéaire à marge dure	Classification supervisée	$\hat{y}_i = \text{sign}(w^T x_i + b)$	$\frac{1}{2} \ w\ ^2$	Standardisation
CSV-C (C-Suport Vector Classification)	Régression fortement non linéaire avec tolérance aux marges (hyperparamètre non borné)	Classification supervisée	$\hat{y}_i = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b)$	$\frac{1}{2} \ w\ ^2 + C \sum_{i=1}^n \xi_i$	Standardisation
ν -SVM (ν -Suport Vector Machine)	Régression fortement non linéaire avec tolérance aux marges (hyperparamètre borné)	Classification supervisée	$\hat{y}_i = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b)$	$\frac{1}{2} \ w\ ^2 - \nu \rho + \frac{1}{n} \sum_{i=1}^n \xi_i$	Standardisation
SVR / ϵ -SVM (Support Vector Regressor)	Régression fortement non linéaire	Régression supervisée	$\hat{y}_i = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b$	$\frac{1}{2} \ w\ ^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$	Standardisation
ν -SVR (ν -Support Vector Regressor)	Régression fortement non linéaire	Régression supervisée	$\hat{y}_i = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b$	$\frac{1}{2} \ w\ ^2 + C(\nu\epsilon + \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*))$	Standardisation

TABLE 3.2 – Synthèse des modèles de Machine Learning - vecteurs support

Modèle	Quand l'utiliser	Paradigme	Formule	Quantité à minimiser	Préparation des données
Decision tree	Classification	Classification supervisée	Partitionnement récursif	$\sum_{i=1}^n (y_i - \hat{y}_i)$ dans les feuilles	-
Random Forest	Classification, robustesse, relations non linéaires	Méthode ensembliste (classification)	Moyenne ou classe majoritaire des arbres	$\sum_{i=1}^n (y_i - \hat{y}_i)$ out-of-bag	-
Gradient Boosting / XGBoost	Classification, robustesse, relations fortement non linéaire	Ensemble boosting	Actualisation d'un arbre faible	$-\sum y \log(\hat{y})$	-

TABLE 3.3 – Synthèse des modèles de Machine Learning - méthodes d'ensemble

Modèle	Quand l'utiliser	Paradigme	Formule	Quantité à minimiser	Préparation des données
AR	Séries temporelles	Modèle autorégressif		-	Stationnarité
MA	Séries temporelles	Modèle autorégressif		-	Stationnarité
ARMA	Séries temporelles	Modèle autorégressif		-	Stationnarité
ARIMA	Séries temporelles	Modèle autorégressif		-	Stationnarité
SARIMA	Séries temporelles, saisonnarité	Modèle autorégressif		-	Stationnarité
SARIMAX	Séries temporelles, saisonnarité avec variables exogènes	Modèle autorégressif		-	Stationnarité

TABLE 3.4 – Synthèse des modèles de Machine Learning - processus autorégressifs

Modèle	Quand l'utiliser	Paradigme	Formule	Quantité à minimiser	Préparation des données
AR	Séries temporelles	Modèle autorégressif		-	Stationnarité

TABLE 3.5 – Synthèse des modèles de Machine Learning - clustering

Modèle	Quand l'utiliser	Paradigme	Formule	Quantité à minimiser	Préparation des données
AR	Séries temporelles	Modèle autorégressif		-	Stationnarité

TABLE 3.6 – Synthèse des modèles de Machine Learning - réduction de dimension

