

Lire les métriques

Une matrice de confusion avant tout score isolé.

Une métrique répond à une question

CLASSIFICATION · LoL

Prédit-on la bonne classe ?
Quels faux positifs / faux négatifs ?

RÉGRESSION · Vin

De combien de points de note se trompe-t-on ?
Les grosses erreurs comptent-elles davantage ?

Fixer la métrique avant de comparer les modèles

Matrice de confusion · 100 parties LoL fictives

		PRÉDICTION DU MODÈLE	
		défaite (0)	victoire (1)
RÉALITÉ blueWins	défaite (0)	VN 45	FP 5
	victoire (1)	FN 10	VP 40

Lignes = vraies classes ; colonnes = classes prédites ; positive = victoire (1).

Quatre calculs à partir de la même matrice

ACCURACY

proportion totale de bonnes prédictions
 $\Rightarrow (VP + VN) / N = 85 / 100 = 0,85$

PRÉCISION

parmi les cas prédits positifs, combien sont vraiment positifs ?
 $\Rightarrow VP / (VP + FP) = 40 / 45 \approx 0,89$

RAPPEL

parmi les vrais positifs, combien le modèle en retrouve ?
 $\Rightarrow VP / (VP + FN) = 40 / 50 = 0,80$

F1

compromis entre précision et rappel. Il est élevé seulement si les deux sont bons.
 $\Rightarrow 2 \times \text{précision} \times \text{rappel} / (\text{précision} + \text{rappel}) \approx 0,84$

Quatre calculs à partir de la même matrice

ACCURACY

À utiliser quand les classes sont équilibrées et que les deux types d'erreur ont à peu près le même coût.

PRÉCISION

À privilégier quand les faux positifs sont coûteux. Par exemple : filtrage spam, alerte fraude, système qui bloque un utilisateur.

RAPPEL

À privilégier quand les faux négatifs sont coûteux, par exemple dépistage médical, détection de fraude ou de panne.

F1

À utiliser quand les classes sont déséquilibrées et qu'on veut éviter d'avoir une bonne précision mais un mauvais rappel, ou l'inverse.

Pourquoi l'accuracy peut masquer le problème

EXEMPLE FICTIF

95 défaites, 5 victoires.
Prédire toujours « défaite » : accuracy = 95 %.

MAIS...

Rappel de la classe « victoire » = 0 %.
Aucune victoire détectée.

- Toujours comparer à une baseline et lire la matrice de confusion
- Choisir la métrique selon le coût des erreurs.

Seuil et AUC : deux questions différentes

SEUIL

Convertit une probabilité ou un score en classe.
Le déplacer change FP, FN, précision et rappel.

AUC ROC

Résume la capacité de classement sur les seuils.
Ne fixe pas le seuil adapté à l'usage.

Ne pas régler un seuil après avoir vu le test final

Régression · l'erreur est une distance

MAE

Moyenne des $|\text{note réelle} - \text{prédite}|$.
En points de note ; moins = mieux.

RMSE

Racine de l'erreur quadratique moyenne.
Pénalise davantage les grosses erreurs.

R^2

Compare aux prédictions constantes de la moyenne.
Peut être négatif sur un jeu évalué.

BASELINE

Toujours comparer à DummyRegressor.
Regarder aussi les erreurs individuelles.

Traduire en appels scikit-learn

```
from sklearn.metrics import (  
    ConfusionMatrixDisplay, accuracy_score,  
    precision_score, recall_score, f1_score,  
    mean_absolute_error  
)  
ConfusionMatrixDisplay.from_predictions(y_test, y_pred)  
print(accuracy_score(y_test, y_pred))  
print(mean_absolute_error(y_test_vin, pred_vin))
```

Pour la CV en régression : `scoring='neg_mean_absolute_error'` (score à maximiser).

Références

Métriques et scoring

https://scikit-learn.org/stable/modules/model_evaluation.html

Convention des axes de la matrice

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

Visualiser une matrice avec scikit-learn

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.ConfusionMatrixDisplay.html>