

4. OVERFITTING

LEV KIWI

APPRENTISSAGE SUPERVISÉ

The diagram shows a table with 5 columns: Name, Balance, Age, Employed, and Write-off. A bracket above the first four columns is labeled 'Attributes', and an arrow points from the 'Write-off' column to the label 'Target attribute'. The row for 'Claudio' is highlighted in blue. An arrow points from this row to a text block below the table.

Name	Balance	Age	Employed	Write-off
Mike	\$200,000	42	no	yes
Mary	\$35,000	33	yes	no
Claudio	\$115,000	40	no	no
Robert	\$29,000	23	yes	yes
Dora	\$72,000	31	no	no

This is one row (example).

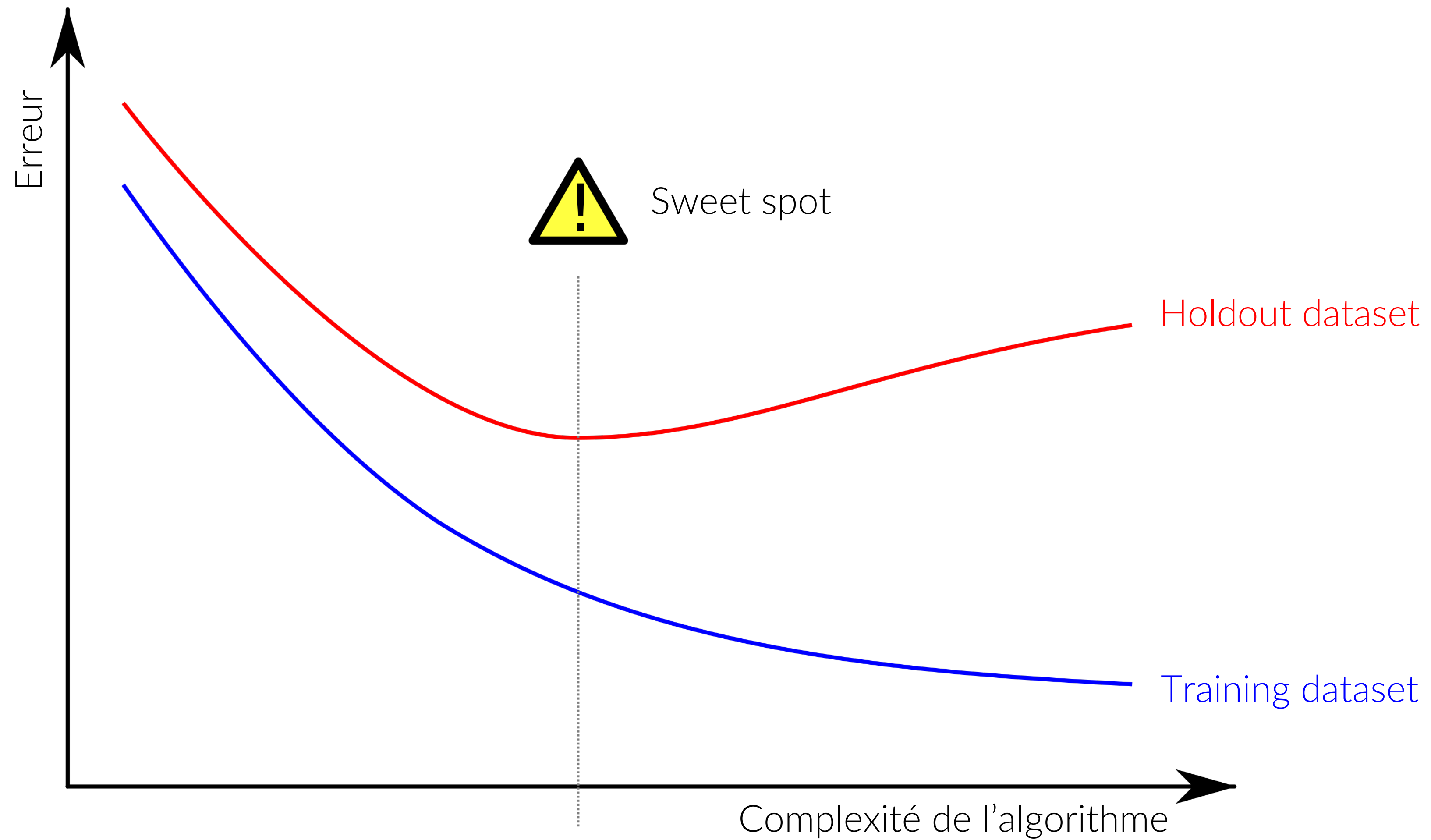
Feature vector is: **<Claudio,115000,40,no>**

Class label (value of Target attribute) is **no**

Dataset d'entraînement

- Les données d'entraînement sont **labélisées**
- L'algorithme va créer des associations entre les **attributs** et la **target**

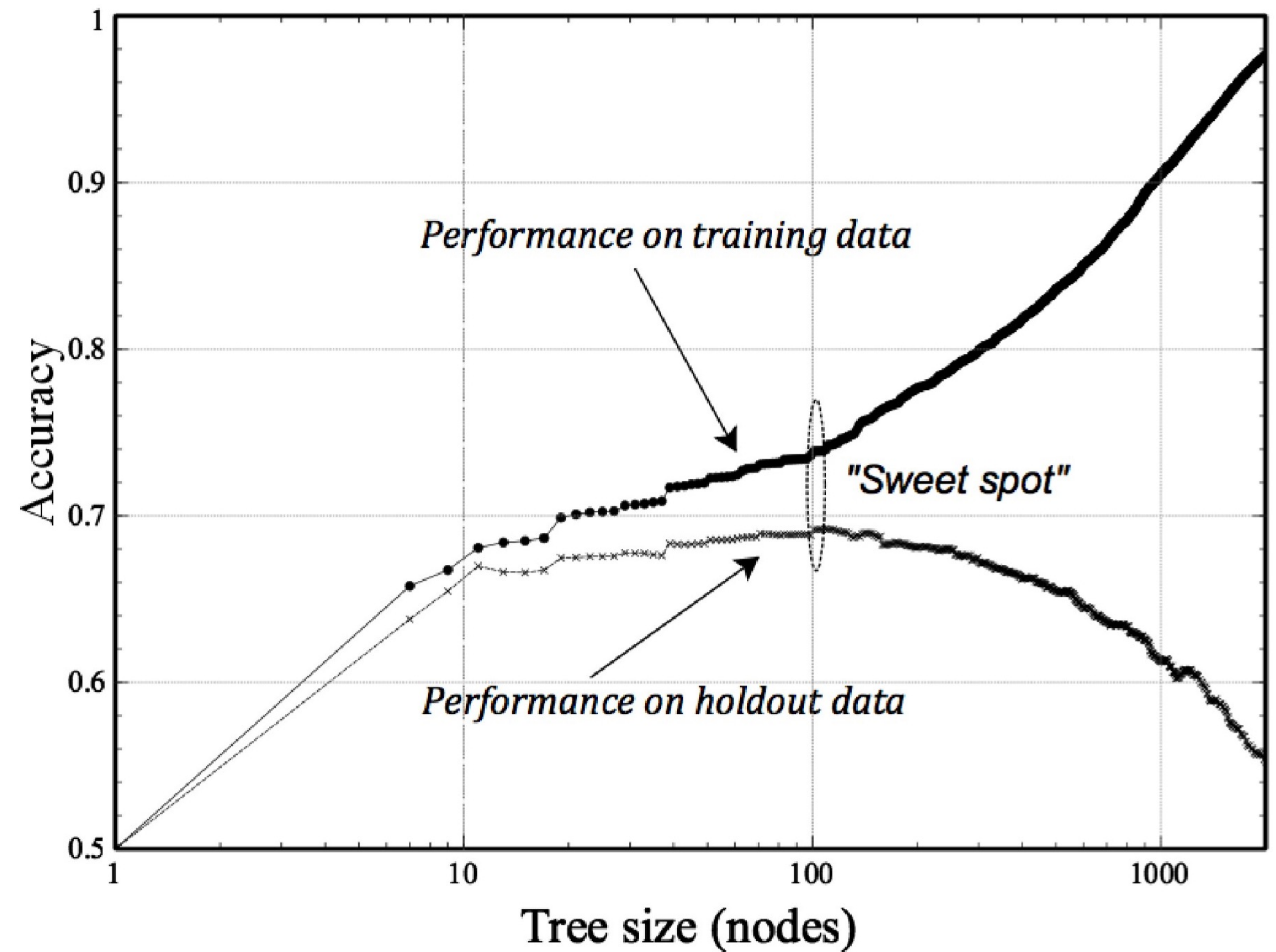
FITTING GRAPH



ARBRE DE DÉCISION

Exemple

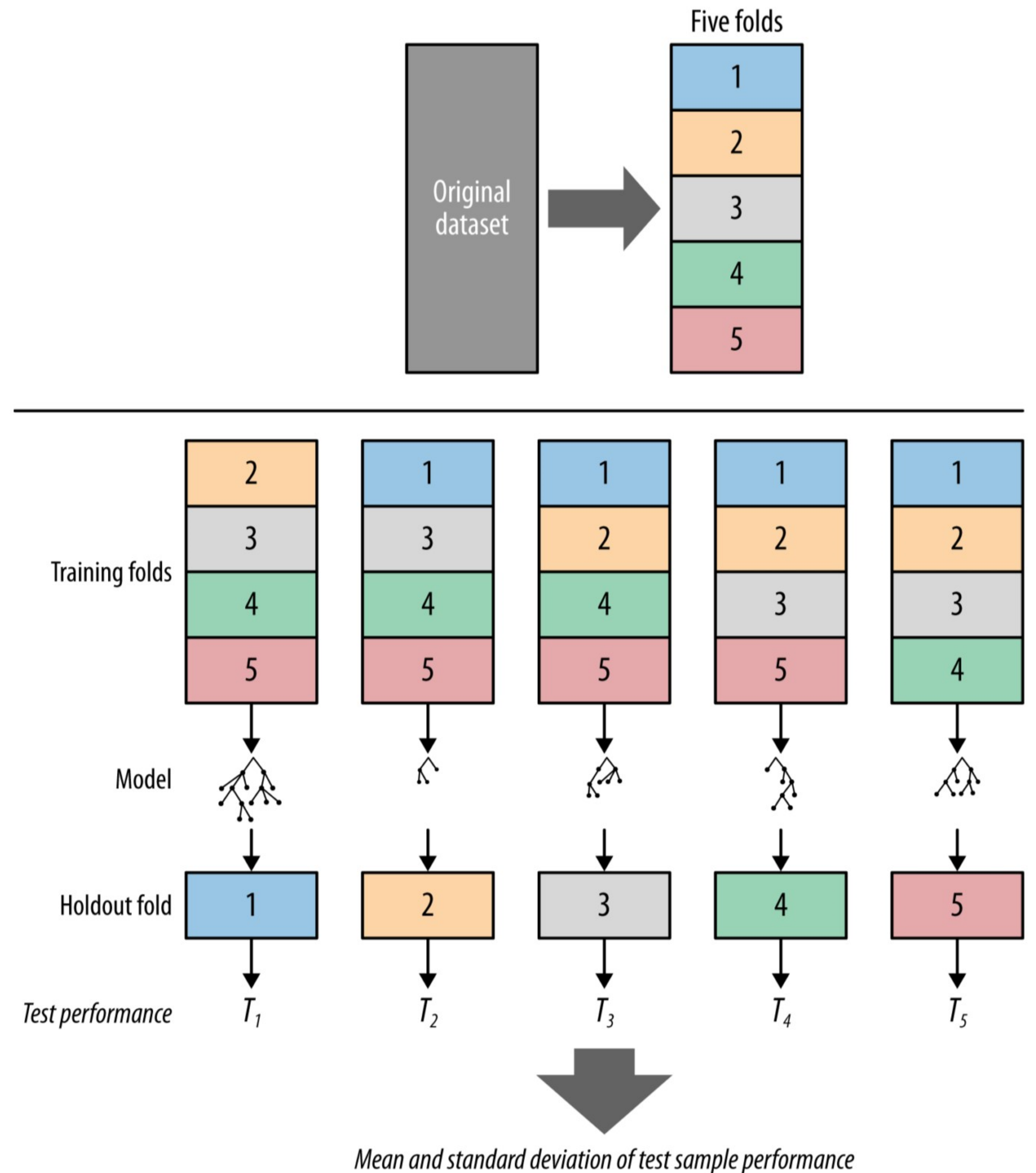
Un arbre de décision est entraîné avec de plus en plus de nœuds.



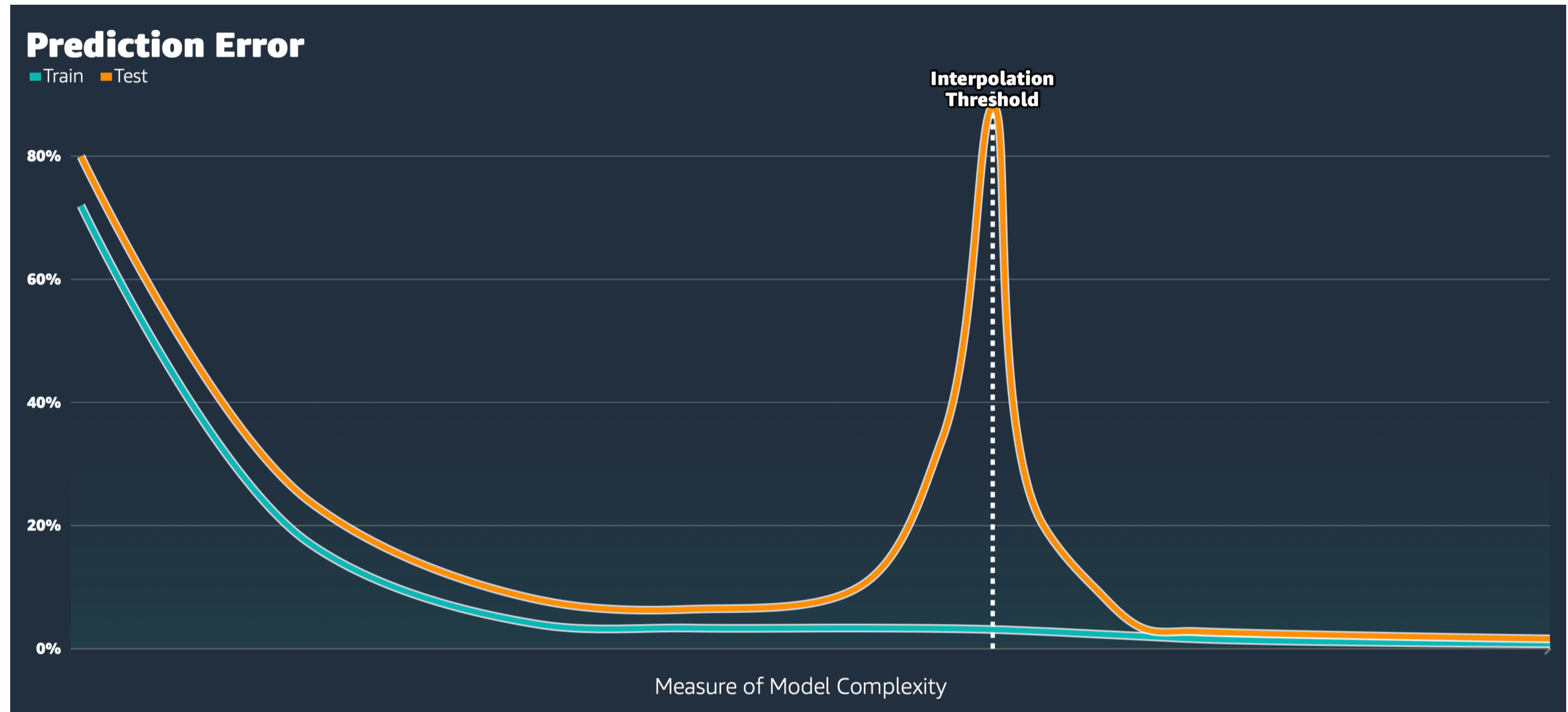
CROSS-VALIDATION

Procédure

1. Le dataset est découpé en k-parties aléatoires de tailles égales.
2. En excluant chaque fois une partie du dataset, le modèle est entraîné sur les k-1 parties restantes.
3. La performance est calculée sur chacune des parties
4. La performance totale du modèle est la performance moyenne.



DOUBLE DESCENT



<https://mlu-explain.github.io/double-descent/>