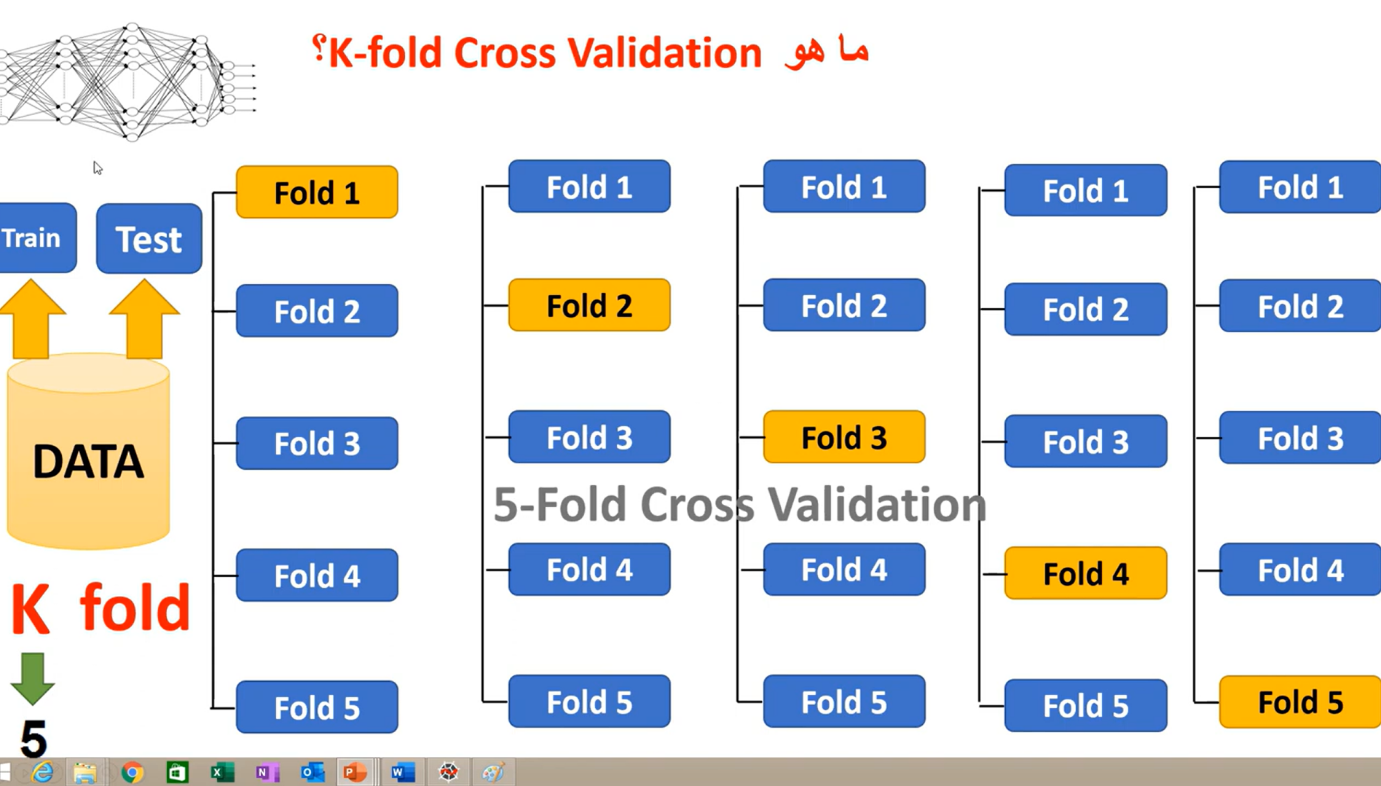
# **Cource-search**

**🡪 k-fold**

**🡪 Over and underfitting**

## **K-Fold Cross-Validation-:**

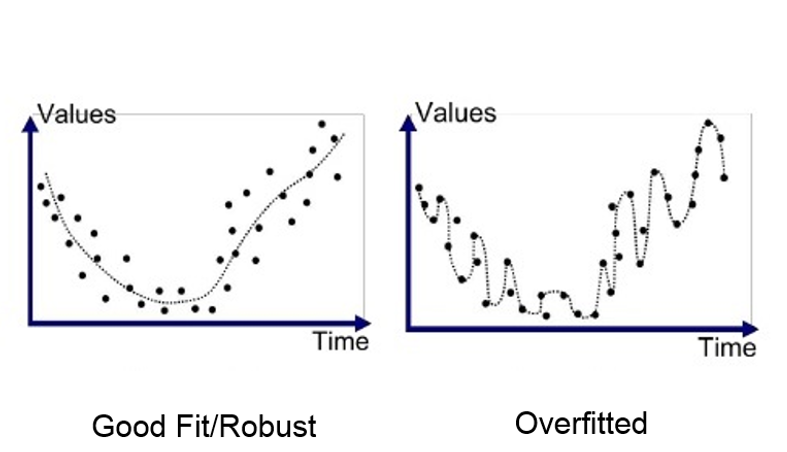
هو تقنية تُستخدم لتقييم أداء نموذج التعلم الآلي بطريقة أكثر دقة وموثوقية. الفكرة الرئيسية هي تقسيم البيانات إلى عدة مجموعات ، واستخدام كل مجموعة كبيانات اختبار مرة واحدة فقط، بينما تُستخدم المجموعات الأخرى لتدريب النموذج وهكذا مع كل مجموعه حيث كل مجموعه لها دورها بأن تكون لاجل ال تيست والباقي لاجل الترين



## Overfitting and underfitting-:

Overfitting and underfitting are common issues in machine learning models. Understanding their causes can help in selecting the right techniques to mitigate them.

Overfitting-->



Overfitting occurs when a model learns the noise and details in the training data to the extent that it performs well on the training data but poorly on new, unseen data. This usually happens due to the following reasons:

Complex Model:

High Capacity: Models with too many parameters (like deep neural networks) can fit almost any function, including noise in the training data.

Low Bias: The model has enough flexibility to capture the peculiarities of the training data, including outliers and noise.

Insufficient Training Data:

When the training dataset is small, the model can easily memorize the training examples rather than learning generalizable patterns.

Too Many Features:

Including a large number of features, especially if many are irrelevant, can cause the model to fit the noise in the training data.

Low Regularization:

Regularization techniques (like L1, L2, dropout) are not used or are too weak, allowing the model to become overly complex and sensitive to the training data.

##################################################

Underfitting -->sA diagram of values and values

Description automatically generated

Underfitting occurs when a model is too simple to capture the underlying structure of the data, leading to poor performance on both the training and test sets. The common reasons for underfitting include:

Simple Model:

Low Capacity: Models with too few parameters (like linear regression

for a non-linear problem) are unable to capture the complexity of

the data.

High Bias: The model is too constrained and cannot learn the underlying patterns in the data.

Insufficient Training Time:

The model has not been trained long enough to learn the underlying patterns, often due to premature stopping in algorithms like gradient descent.

Too Few Features:

Key features that carry important information are missing, preventing the model from capturing the data's complexity.

Over-Regularization:

Excessive use of regularization techniques can overly penalize the model's complexity, resulting in a model that is too simplistic.

Inappropriate Model Selection:

Choosing a model that is not suitable for the data can lead to underfitting. For instance, using a linear model for data with a complex, non-linear relationship.

For Overfitting:

Simplify the model by reducing the number of parameters.

Use regularization techniques like L1, L2 regularization, or dropout.

Increase the amount of training data.

Perform feature selection to remove irrelevant features.

Use cross-validation to ensure the model generalizes well.

For Underfitting:

Increase the model complexity by adding more parameters.

Reduce the amount of regularization.

Add more relevant features.

Train the model for a longer period or with more advanced optimization techniques.

Choose a more suitable model that matches the data complexity.

##########

Understanding and addressing overfitting and underfitting is crucial for building robust machine learning models that generalize well to new data.