Cross-validation

Cross-validation in **Machine Learning (ML)** is a **model evaluation technique** used to test how well a model generalizes to unseen data. Instead of training and testing on a single split, cross-validation splits the dataset into multiple parts (called **folds**) and trains/tests the model multiple times for a more reliable performance estimate.

🔑 **Why Cross-Validation?**

•Prevents **overfitting** (model doing well on training data but poorly on new data).

•Provides a **better estimate** of model performance than a single train/test split.

•Ensures that **all data points are used** for both training and testing at some stage
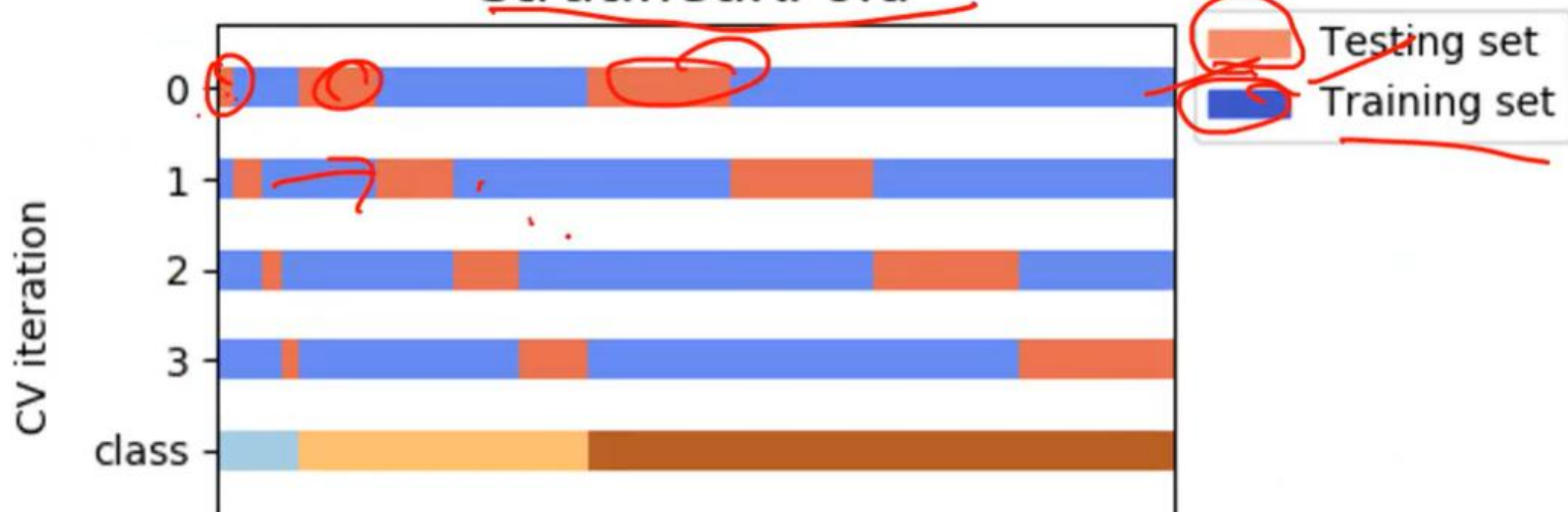
Cross Validation

StratifiedKFold

🏷️ **Types of Cross-Validation**

1.K-Fold Cross-Validation

    1. Data split into **k folds** (commonly k=5 or k=10).

    2. Model trained & tested k times, results averaged.

2.Stratified K-Fold

    1. Ensures each fold has the **same proportion of classes** as the full dataset.

    2. Useful in **classification with imbalanced data**.

3.Leave-One-Out Cross-Validation (LOOCV)

    1. Special case of k-fold where **k = number of samples**.

    2. Train on all data except 1 point, test on that 1.

    3. Very accurate but **computationally expensive**.

4.Repeated K-Fold

    1. K-fold repeated multiple times with different random splits.

    2. Gives a **more stable estimate**.

5.Time Series Cross-Validation

    1. Used when data has a **temporal order** (e.g., stock prices, weather).

    2. Training only uses **past data** to predict **future data**.