

Transversal Machine Learning Concepts

Baseline, Validation and Evaluation

Kheireddin Kadri

DVRC-Simplon-PSTB

October 20, 2025

Outline

- 1 Baseline Model
- 2 Splitting the Dataset
- 3 Learning Curves
- 4 Cross-Validation
- 5 Grid / Random Search
- 6 Classification Metrics

Baseline Model

Definition

A **baseline model** is a simple reference point used to verify that a more complex model actually improves performance.

Examples

- Classification – always predict the majority class.
- Regression – predict the mean or median of the target.

Why it matters

If your model does not beat the baseline, it has learned nothing useful.

Train / Validation / Test Split

Principle

Divide the dataset to evaluate how well the model generalizes.

- Train – fit the parameters.
- Validation – tune hyperparameters.
- Test – unbiased final check.

Python

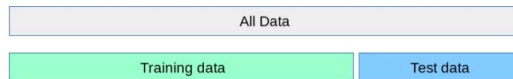


Figure: Sample a training set 80% while holding out 20% of the data for testing.
(https://scikit-learn.org/stable/modules/cross_validation)

Learning Curves

Purpose

Show how training and validation scores evolve with the number of samples.

Reading the curves

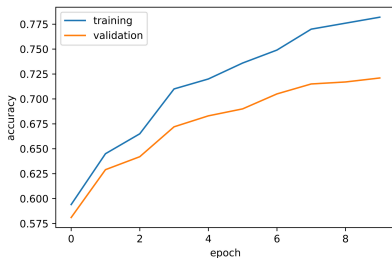
- Large gap \rightarrow overfitting.
- Both low \rightarrow underfitting.
- Converge high \rightarrow good fit.

Learning Curves: Loss and Accuracy

Visualization

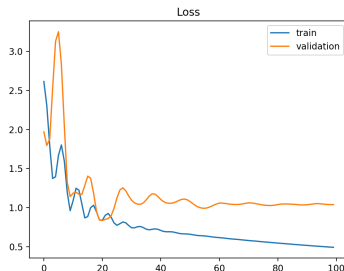
Learning curves help us diagnose bias/variance trade-offs by observing how **accuracy** and **loss** evolve for both training and validation sets.

Accuracy Curves



Train vs Validation Accuracy

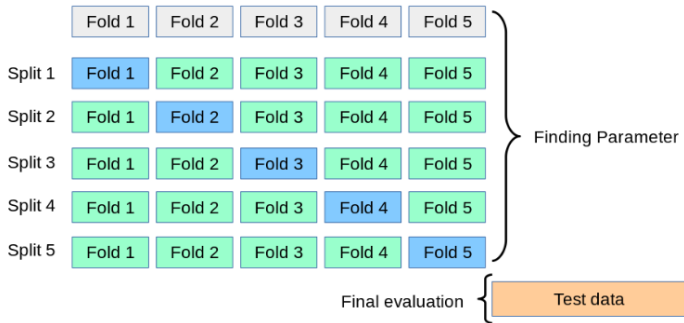
Loss Curves



Train vs Validation Loss

Idea

Split data into k folds; train on $k - 1$, test on the remaining one, repeat k times.



k -fold CV, the training set is split into k smaller sets.

Hyperparameter Tuning

Goal

Find the best combination of hyperparameters.

Approaches

Grid Search – exhaustive combinations.

Random Search – random samples in parameter space.

Python

Main Metrics

- Accuracy = correct / total
- Precision = $TP / (TP + FP)$
- Recall = $TP / (TP + FN)$
- F1 = $2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall})$

Confusion Matrix

Summarizes TP, FP, TN, FN.

Python

Checklist

- Build a baseline.
- Split properly.
- Plot learning curves.
- Validate with CV.
- Tune with Grid/Random Search.
- Evaluate with multiple metrics.