[SWCON253] Machine Learning – Lec.13

Model Evaluation

Fall 2025

김휘용

hykim.v@khu.ac.kr

경희대학교
KYUNG HEE UNIVERSITY

Contents

1. Introduction
2. Training, Test, Validation Sets
3. Holdout Methods
4. Cross-Validation

References

- *Intro to Machine Learning* by Sebastian Raschka (http://stat.wisc.edu/~sraschka/teaching/stat451-fs2020/)

Introduction

Model Evaluation

- Estimates the ***generalization performance***, i.e., the predictive performance of our model on future (*unseen*) *data*.

1. ***Model Test***: Want to ***evaluate*** the generalization performance of the ***developed (final) model*** before deploy.

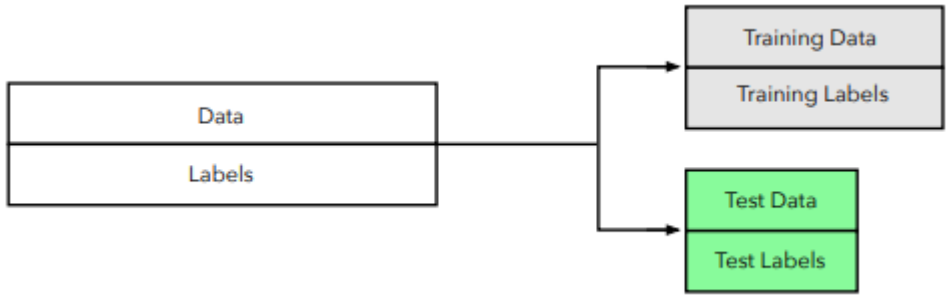
2. ***Model Validation***: Want to ***increase*** the generalization performance ***in the development stage*** by tweaking the learning algorithm and *selecting the best* performing model from a given hypothesis space.

Training, Test, Validation Sets

- **Training Set**
  - Use training set ***to train*** your model (***model fitting***)
- **Test Set**
  - Use test set ***only to evaluate*** the performance of your final model (***model test***)
  - *Caution: Do not use test set (or test result) to modify your model!*
- **Validation Set**
  - You can use validation set ***to select*** *best* model or hyper-parameters (***model selection***)
    - ★ e.g., in polynomial regression, select the order of the polynomial (i.e., model capacity)
    - ★ e.g., in L1 or L2 regularization, selecting the regularization-weight λ
    - ★ e.g., in SVM, selecting the slack-weight C and/or kernel parameters
    - ★ Note: "*best*" here means best performing on the validation set, implying the lowest generalization error

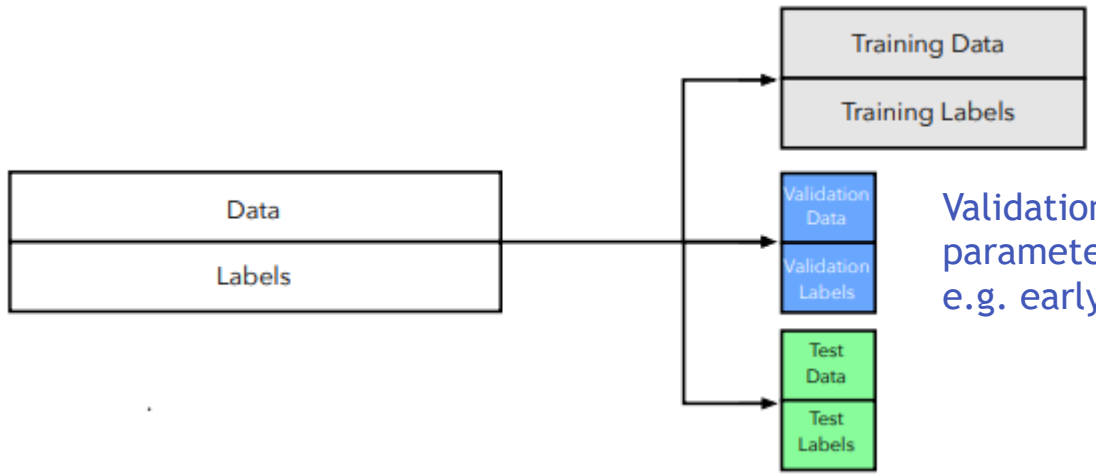
2-Way Holdout Method

- **The Simplest Technique for Testing a Model**
  - ① *Divide* our available data into two subsets: a ***training*** and a ***test*** set.
    - ★ e.g., randomly assign 33% of the data to the test set, or assign 10% if the dataset is relatively large.
  - ② *Train* your ML algorithm with the training set.
  - ③ *Estimate* the generalization *performance* of the trained model with the test set.
  - ④ (*Optional*) *Retrain* your ML algorithm with *all* the data, including both the training set and test set.
    - ★ As a rule of thumb, the model will have a better generalization performance if the algorithms uses more informative data.
    - ★ *Do not re-estimate the generalization performance with this retrained model!*



3-Way Holdout Method

- **Holdout Method for Selecting (Validating) a Model**
  - ① *Divide* our available data into three subsets: a ***training***, a ***validation***, and a ***test*** set.
    - ★ e.g., assign 33% of the data to the test set, or assign 10% if the dataset is relatively large.
  - ② *Train* your ML algorithm with the same training set but *with different hyperparameter settings*.
  - ③ *Compare* the generalization *performance* of the trained models with the *validation* set and *Choose* the *hyperparameters* settings associated with the best performance.
  - ④ *Estimate* the generalization *performance* of the trained model with the *test* set.
  - ⑤ (*Optional*) *Retrain* your ML algorithm with *all* the data, i.e., including both the training set and test set.



Validation Set is extensively used to determine (selecting) model parameters across multiple iterations (epochs)
e.g. early stopping, hyper-parameter tuning

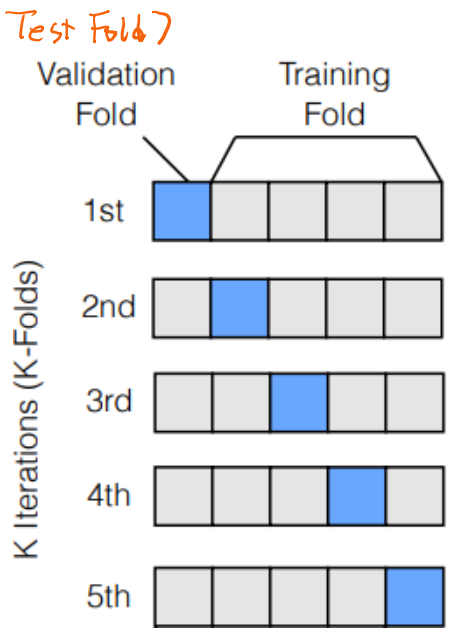
Cross-Validation

Cross-Validation (CV)

- Cross-validation is very useful *when the available dataset size is small*.
- The main idea is that *each sample in our dataset* has the opportunity of being *tested*.
- Cf.) The term cross-validation is used loosely in literature, where the train/test holdout method is sometimes referred to as a cross-validation technique.

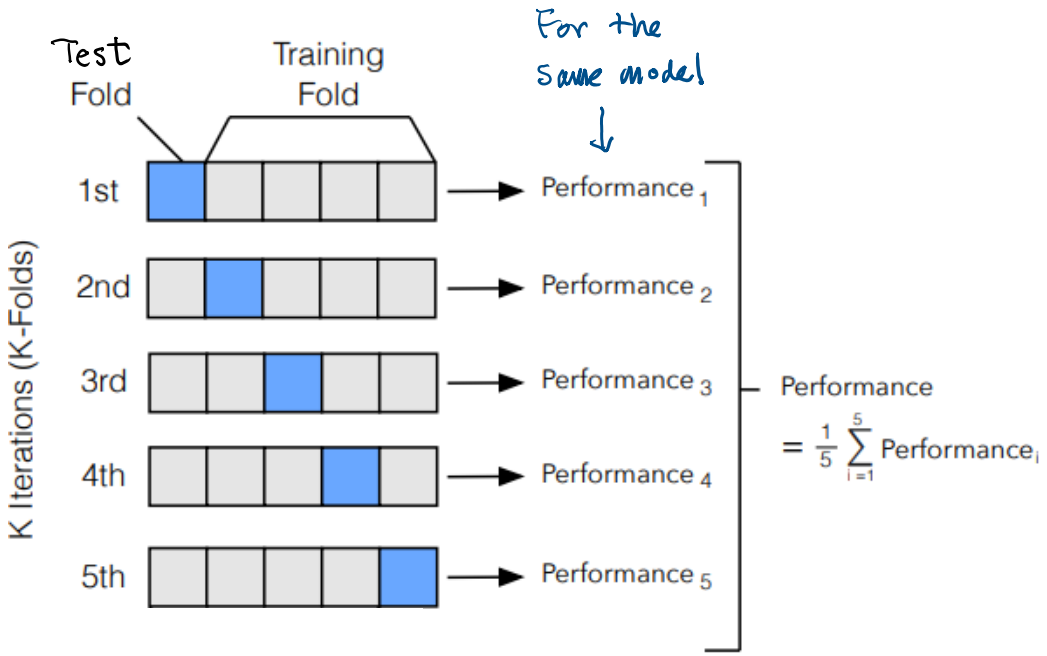
k-Fold Cross-Validation

- ***k-fold cross-validation*** is a special case of cross-validation where we *iterate over a dataset set k times*.
  - ★ non-overlapping test (validation) folds; ***utilizes all data for testing*** (validation)
  - ★ overlapping training folds
- It is the most common technique for model evaluation & selection

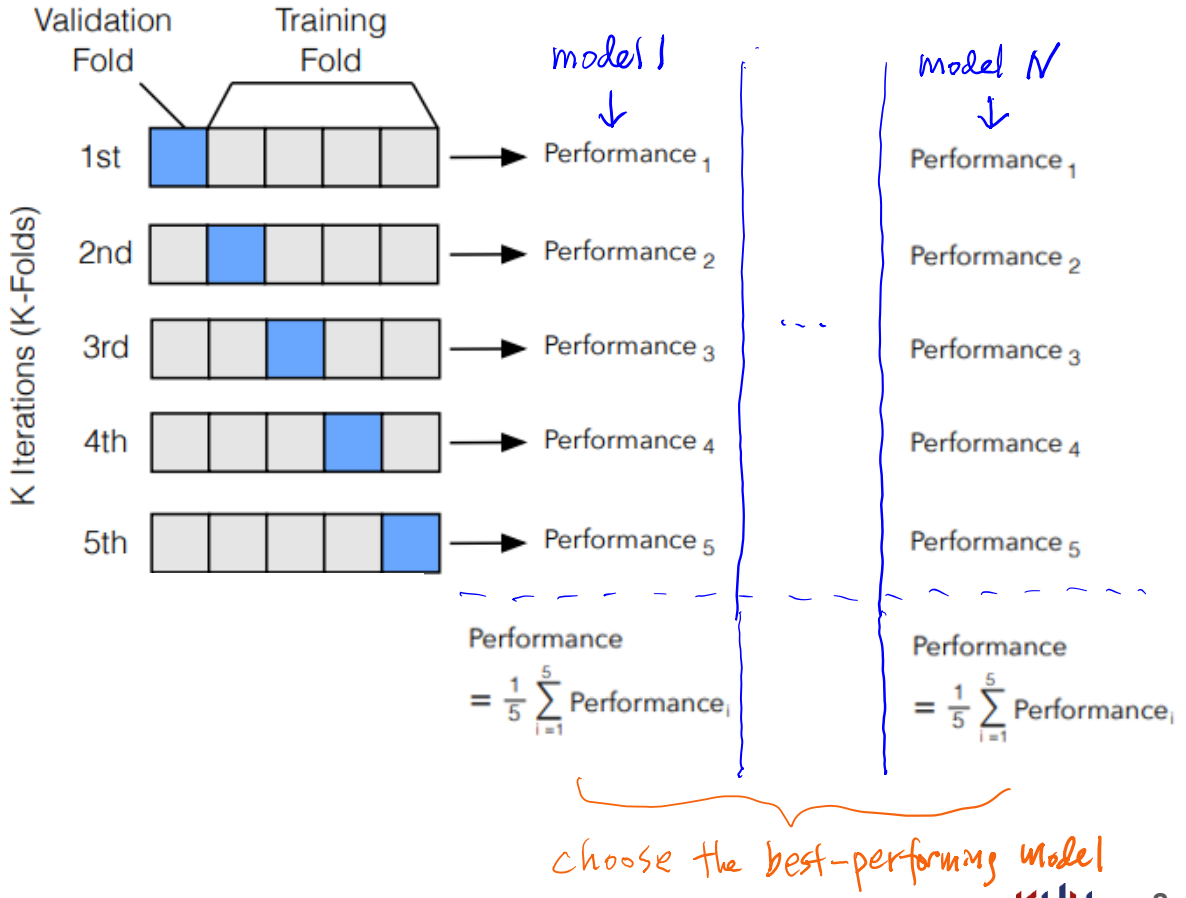


k-Fold Cross-Validation

- ***k*-Fold CV for Model *Testing***
  - Evaluate the generalization performance of a given model (with *fixed hyperparameters*).



- ***k*-Fold CV for Model (or Algorithm) *Selection***
  - Compare the generalization performance of different models (with *different hyperparameters*).
  - Choose the best performing model.



Q & A