Spring 2023

SWCON253: Machine Learning

Lecture 10

Model Evaluation

Jinwoo Choi
Assistant Professor
CSE, Kyung Hee University

Disclaimer: Many slides have been borrowed from Hui Yong Kim, who may have borrowed some of them from others. Any time a slide does not show any explicit credit, I credit the slide to Hui Yong Kim.

Contents

1. Model Evaluation
2. Training, Test, Validation Sets
3. Holdout Methods (2-Way & 3-Ways)
4. k-Fold Cross-Validation

References

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

Model Evaluation

Model Evaluation

- Evaluates the prediction performance of a model
- More specifically, ***estimates*** the ***generalization performance***, i.e., the predictive performance of our model on future (*unseen*) *data*.

Application

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

Test data를 쓰면 안되나

Train에서 추출

(2-Way) Holdout Method

◆ The Simplest Model Test Technique

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.

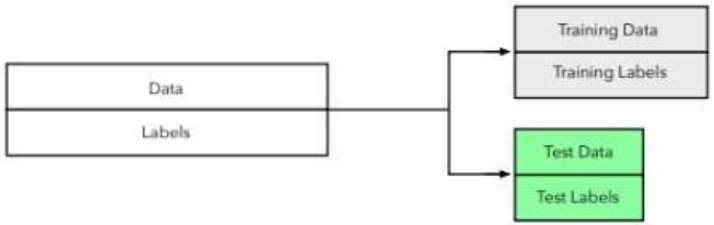
② *Train* your ML algorithm with the training set.

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!*



Test data를 그냥 Validation인지마냥 씀

3-Way Holdout Method

◆ Holdout Method for Model Selection (or Hyperparameter Tuning)

① *Divide* our available data into three subsets: a ***training***, a ***validation***, and a ***test*** set.

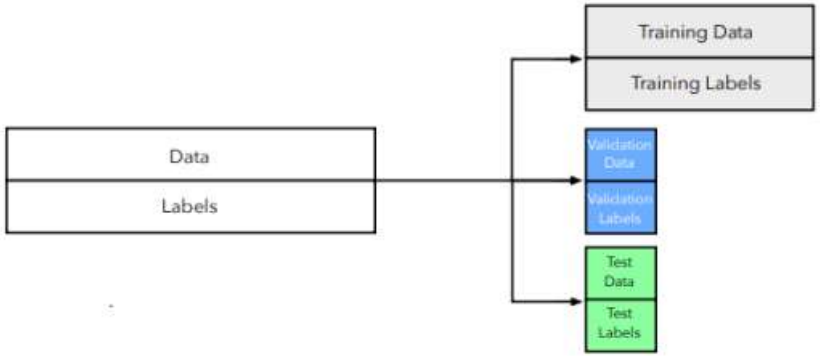
★ e.g., assign 33% of the data to the test set, or assign 10% if the dataset is relatively large.

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.

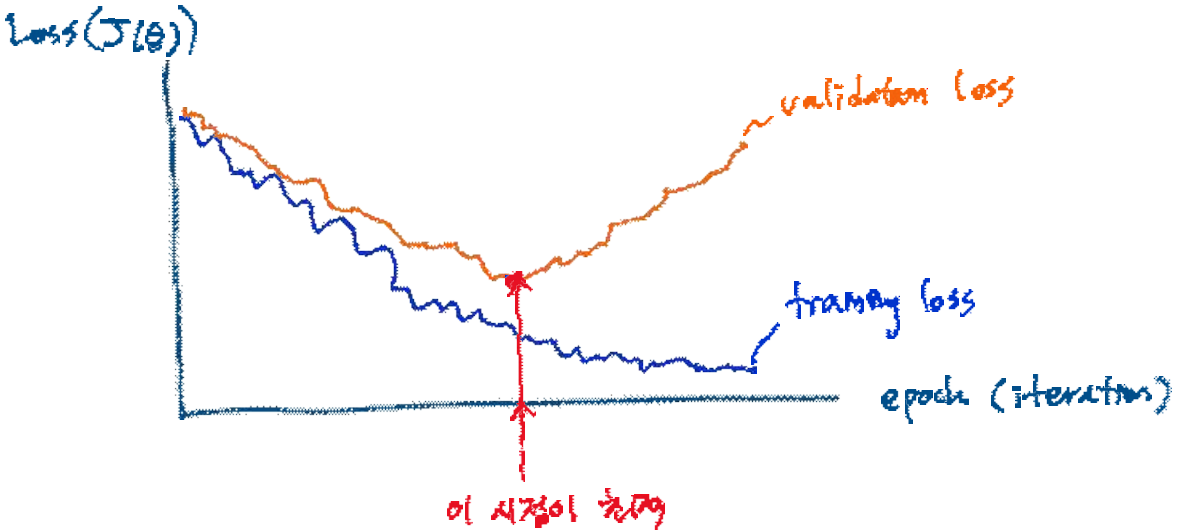
⑤ (*Optional*) *Retrain* your ML algorithm with *all* the data, i.e., including both the training set and test set.



3-Way Holdout Method – *Another Usage*

Early Stopping to Prevent Overfitting

- Validation Set is extensively used to determine (selecting) model parameters across multiple iterations (epochs)
- A kind of the "3-Way Hold-out Method".



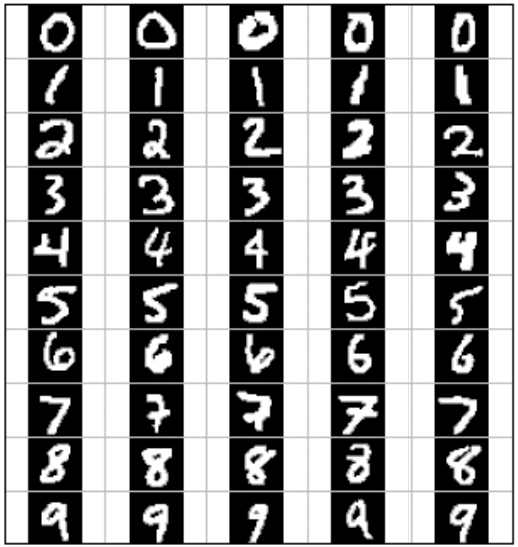
Cross-Validation

Cross-Validation (CV)

- The main idea is that *each sample in our dataset* has the opportunity of being *tested*.
- Cross-validation is very useful when the available *dataset size is small*.
- 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.

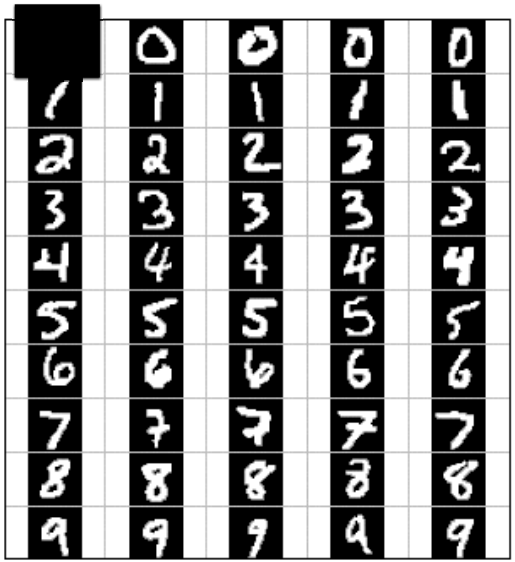
Leave-one-out cross-validation

Held-out Validation



Slide Credit: Bert Huang

Leave-one-out Cross Validation

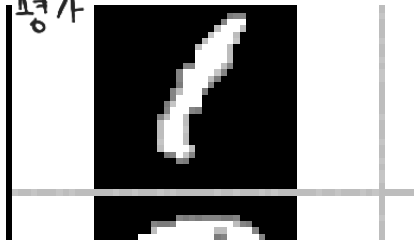


training data

하나씩 빼서 나머지 샘플들로 평가

첫 0 빼고 평가

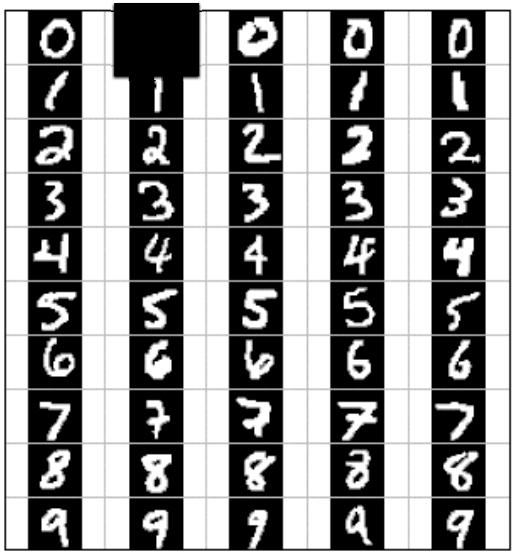
다음 //



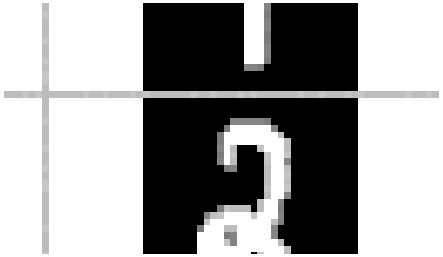
validation data

Slide Credit: Bert Huang

Leave-one-out Cross Validation



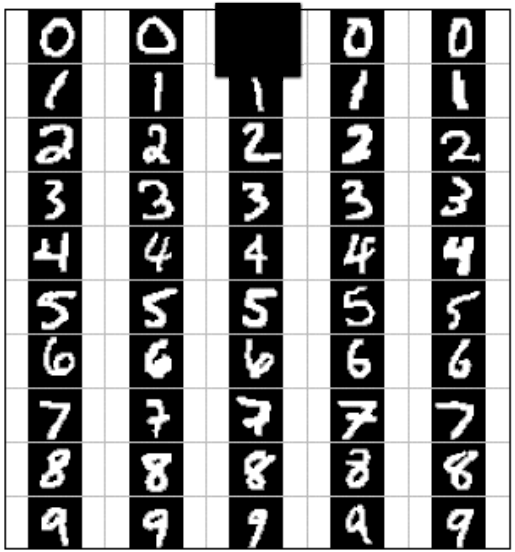
training data



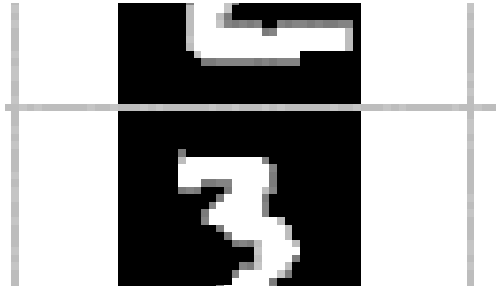
validation data

Slide Credit: Bert Huang

Leave-one-out Cross Validation



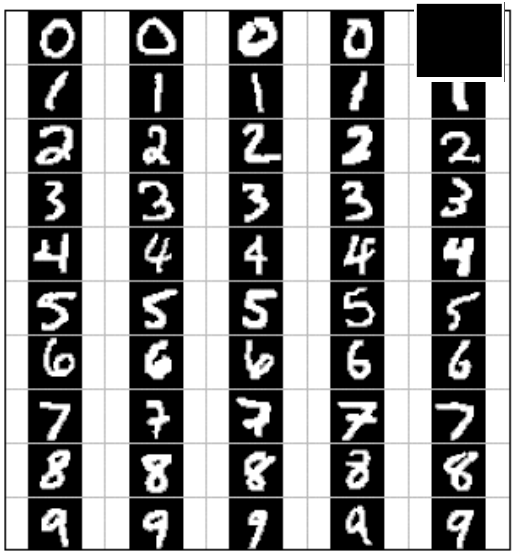
training data



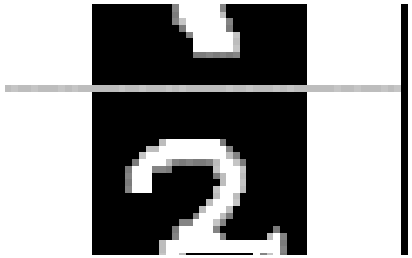
validation data

Slide Credit: Bert Huang

Leave-one-out Cross Validation



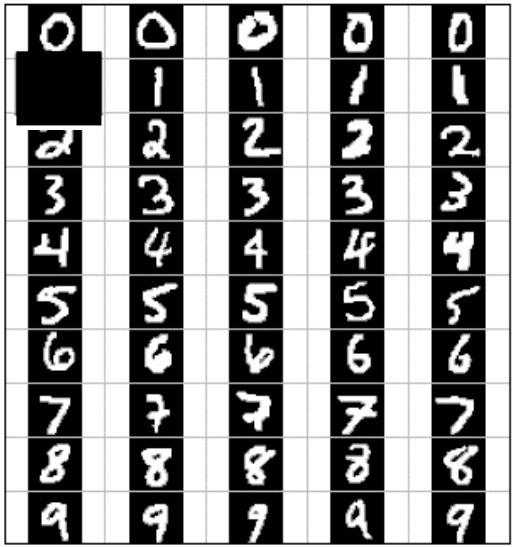
training data



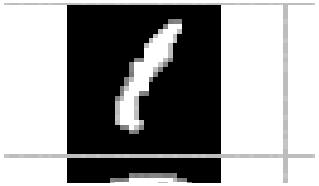
validation data

Slide Credit: Bert Huang

Leave-one-out Cross Validation



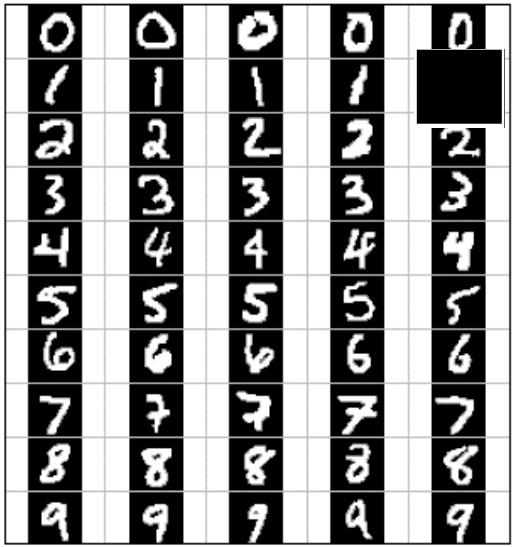
training data



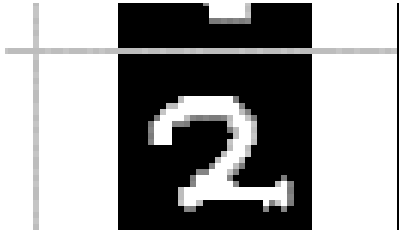
validation data

Slide Credit: Bert Huang

Leave-one-out Cross Validation



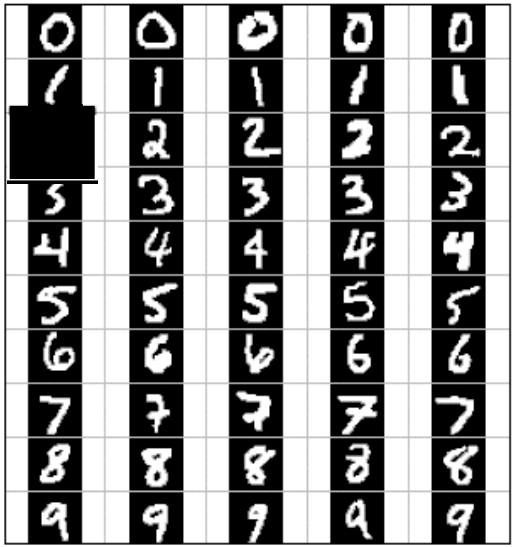
training data



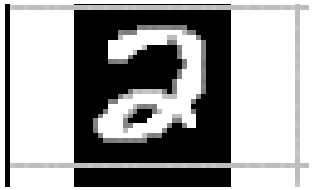
validation data

Slide Credit: Bert Huang

Leave-one-out Cross Validation



training data



validation data

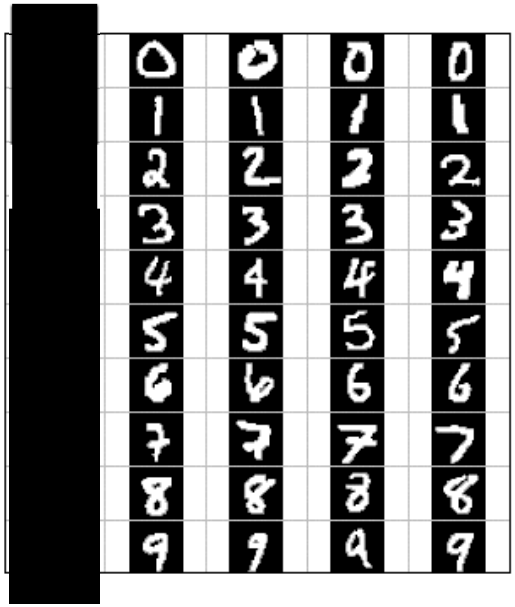
Slide Credit: Bert Huang

Cross-Validation

- **Cross-Validation (CV)**
  - The main idea is that *each sample in our dataset* has the opportunity of being *tested*.
  - Cross-validation is very useful when the available *dataset size is small*.
  - 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.
- ***Leave-one-out cross-validation***
- ***k*-Fold Cross-Validation**

Cross Validation K겹 교차 검증

Fold 1



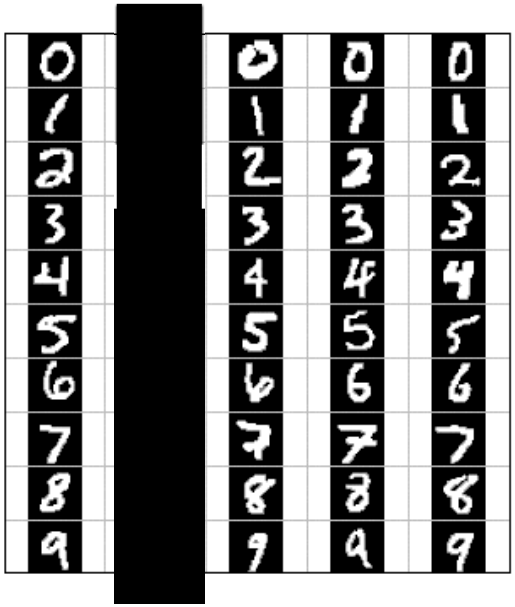
training data

validation data

Slide Credit: Bert Huang

Cross Validation

Fold 2



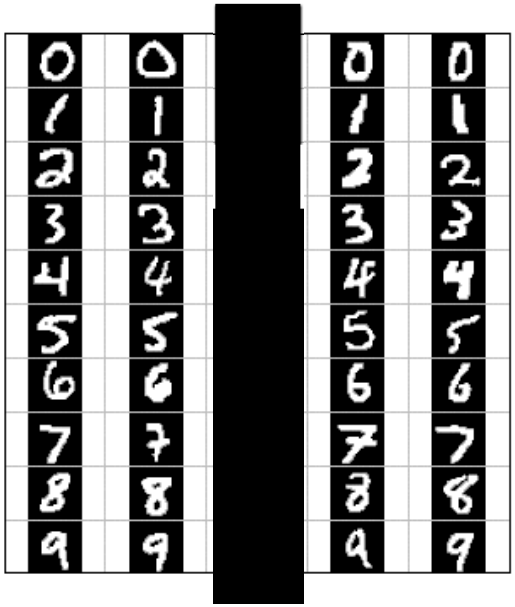
training data

validation data

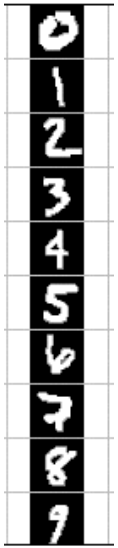
Slide Credit: Bert Huang

Cross Validation

Fold 3



training data

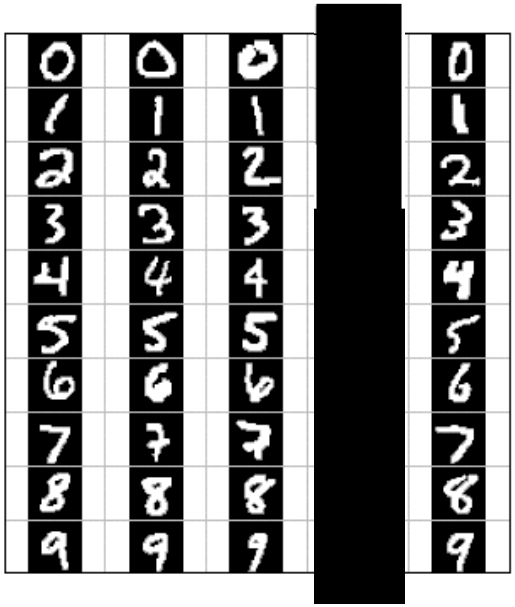


validation data

Slide Credit: Bert Huang

Cross Validation

Fold 4



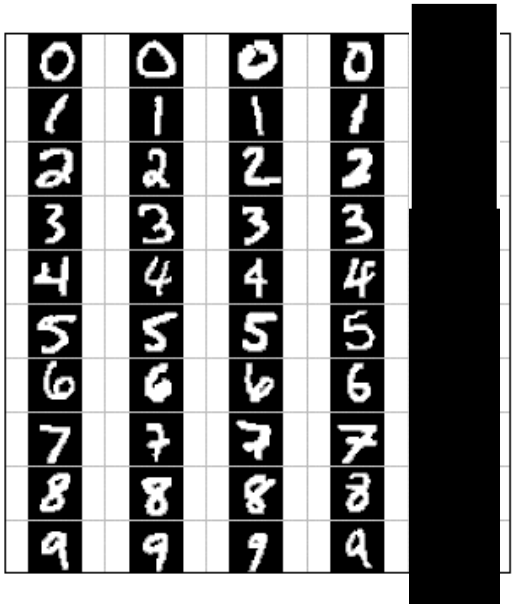
training data

validation data

Slide Credit: Bert Huang

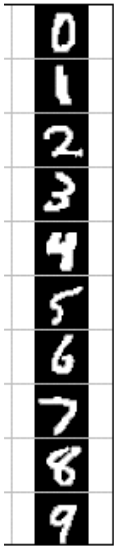
Cross Validation

Fold 5



training data

→5겹 교차 검증



validation data

Slide Credit: Bert Huang

Cross-Validation

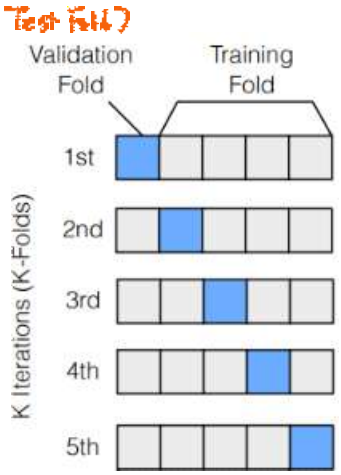
Cross-Validation (CV)

- [illegible] *each sample in our dataset* has the opportunity of being *tested*.
- Cross-validation is very useful when the available *dataset size is small*.
- 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.

Leave-one-out cross-validation

k-Fold Cross-Validation

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



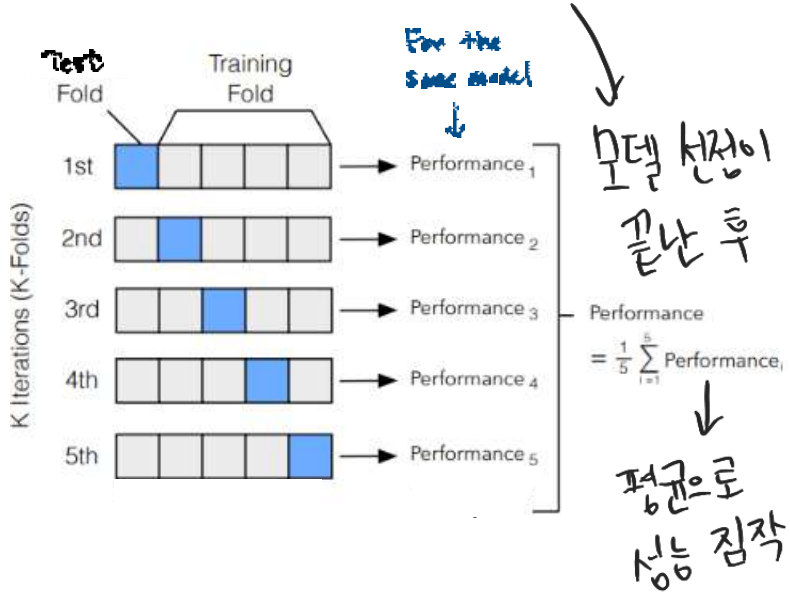
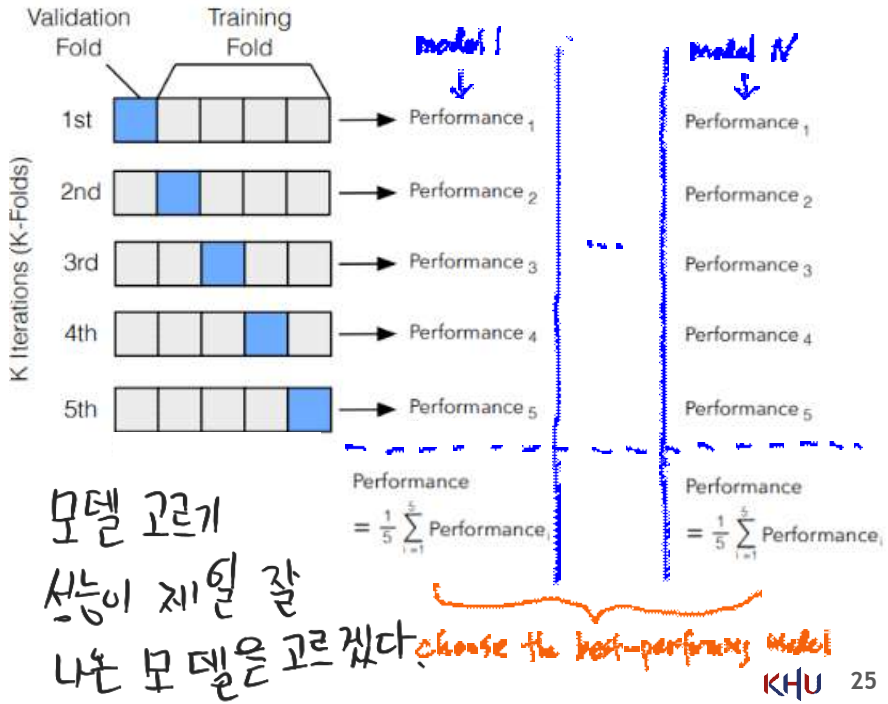
Model Selection via Validation

- Measure performance on **held-out** training data
  - Simulate testing environment
- Rotate **folds** of held-out subsets
- Can even hold out one at a time: **leave-one-out** validation
- Use (cross) validation performance to tune extra parameters

Slide Credit: Bert Huang

k-Fold Cross-Validation

- ***k*-Fold CV for Model *Testing***
  - [illegible] 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.