Artificial Intelligence and Machine Learning  
- Overview -  
Hyeoncheol Kim

## Two Approaches

- **Knowledge-Based Approach**  
  - Symbolic AI, Traditional AI

- **Data-Driven Approach**  
  - Computational AI,  
  - Machine Learning

- **Hybrid Approach**  
  - Knowledge-based + Data-Driven  
    - Ex) Knowledge-based Neural Network  
    - Ex) Knowledge Extraction from Neural Network

## What is that for?

- In order to build a domain model.

### MODEL

- Classification, Prediction, Clustering, Explanation,

## What is the ML for?

- In order to build a domain model.

### MODEL

- Classifier, Predictor, Describer

“Machine Learning” creates a model from a set of data instances  
(e.g., observed data, experience data)

## Model Creation

- **Knowledge-based approach**  
  - Never seen cats or dogs.  
  - Somebody gave you the explicit knowledge of what a cat or a dog is.  
  - You have the knowledge-based model to be used for classification or prediction.

- **Data-driven Approach**  
  - Never been told about the cat or dog concept.  
  - You’ve seen lots of cases of cats and dogs.  
  - You have an implicit concept model to be used for classification or prediction.  
  - Can you explain the concept in explicit format?

## Example: A Domain of “Dog and Cat”

**Question:** Is this a Cat or a Dog?  

To answer this question, we use a so-called “Concept Model (or Concept Knowledge)” built in our brains. We have the “concept model” on a CAT and a DOG already built in our brain. Why can’t we answer the question?
“Machine Learning” creates a model from a set of data instances (e.g., observed data, experience data).

**A Huge Number of Instances**  
- An instance is just one observed example.  
- An instance includes:  
  - Input attributes (or features) with values  
  - Output attribute with a class

**A Model**  
- A model is a concise and generalized form representing the whole domain instances.  
- We can use the model to:  
  - Classify or predict a new instance in the future  
  - Understand the structure of domain information

Examples of a Model:  
- An equation
- A set of simple rules
- A decision tree

**Definition of Machine Learning**

- **“Machine Learning is the study of computer programs that improve automatically through experience”**  

- **ML Algorithms:**  
  - Decision Tree, Neural Networks, HMM, Support Vector Machine, etc.

**Learn What?**

- Class membership  
  – eg. Globins
- Classification hierarchy  
  – eg. Folds, super-families, families
- Sequence prediction  
  – eg. Repeating secondary structure
- Shape descriptions  
  – eg. Binding site descriptions
- Temporal models  
  – eg. Models of cell activity, pathways, etc.

**Training and Testing**

- Domain Dataset
- Training set  
  - To Build a model
- Testing set  
  - To Test model performance

**N-fold Cross-Validation**

- Use (N-1) subsets for training and 1 subset for testing.
- Repeat it N times with different testing subset and Average them.
- General rule for the N:  
  - 10-fold when number of examples, 30≤n<200  
  - Leave-one-out when number of examples, n<30
Training Dataset

- should be domain-representative
- should be class-balanced
- Noise-free

Performance Evaluation

- Accuracy = (TN+TP)/(TP+TN+FP+FN)
- Specificity = TN/(TN+FP)
- Sensitivity = TP/(TP+FN)
- Positive Predictive Value = TP/(TP+FP)
- Negative Predictive Value = TN/(TN+FN)
- Correlation Coefficient = (TP*TN – FP*FN)/SQRT((TP+FP)*(FP+TN)*(TN+FN)*(FN+TP))

Evaluation Measures

- Sensitivity
  - The proportion of all true positive patterns that are correctly identified
- Specificity
  - The percentage of all true negative patterns that are correctly identified
- Positive (or Negative) Predictive Value
  - The probability that a predicted true (or negative) pattern is indeed a true (or negative) pattern
- Correlation coefficient
  - Its value of 1 and -1 correspond to a perfect and a completely wrong prediction, respectively.

Hybrid

- How do we (i.e., our brain) learn?
  - Knowledge-based?
  - Data-driven?
  - Or Both? Incorporated?

Knowledge

- In explicit form
  - Description/Understanding what has been learned
  - Improves learning efficiency

Cognitive concept

- In implicit form
  - Knowledge

- Increases learning efficiency