Ensemble in Machine Learning

Sumahn Cha

Contents

1. VotingClassifier
   - Hard voting
   - Soft voting

2. Stacking
   - General Stacking
   - CV Stacking

3. Multi-class Classification
   - OvO Classifier
   - OvR Classifier

Ensemble Learning

In statistics and machine learning, ***ensemble methods*** *use* ***multiple learning algorithms*** *to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. (Wikipedia)*

How? → Voting
Stacking

Voting

Hard voting

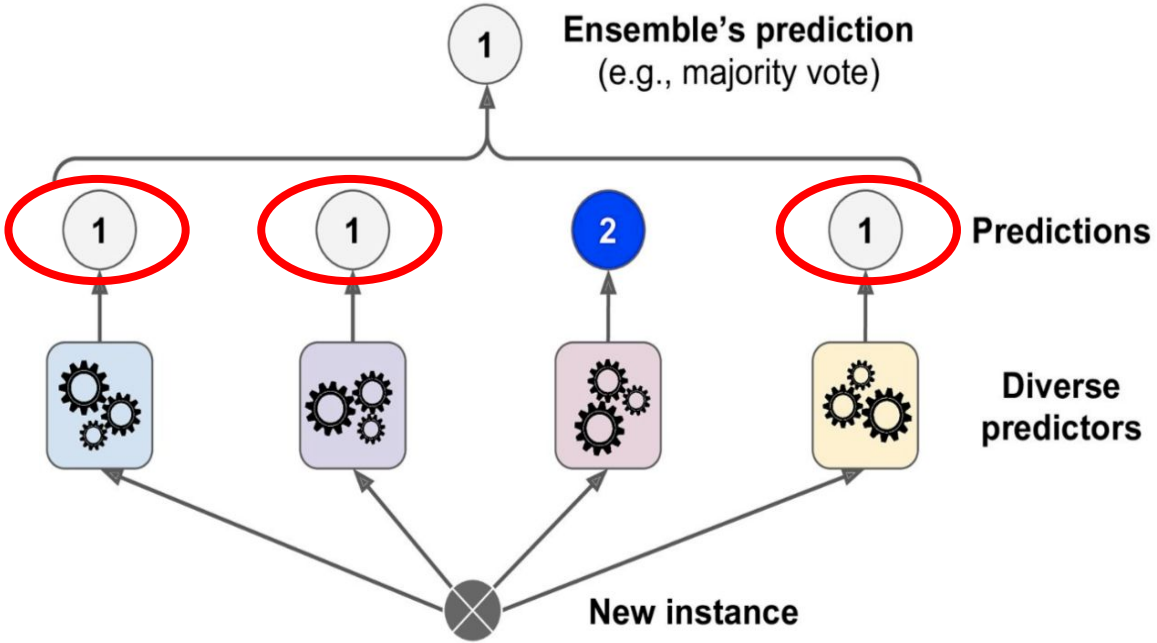
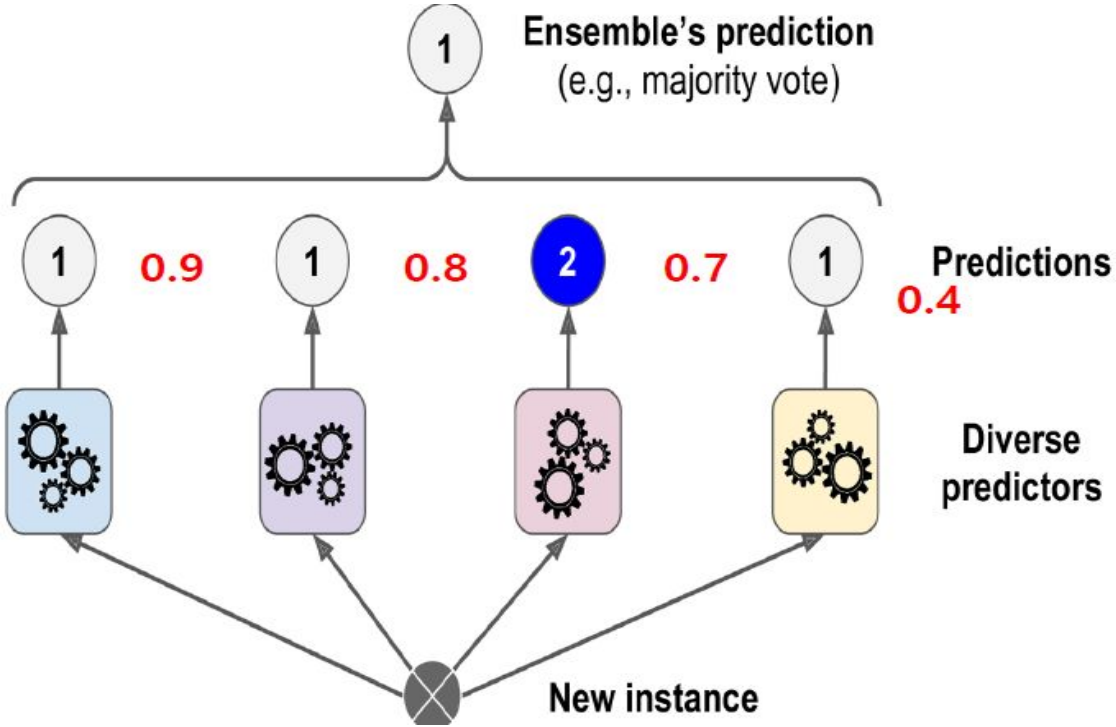


Figure 7-2. Hard voting classifier predictions

1. Based on the results of each classifier
2. Determine the most voted prediction as final result

Voting

Soft voting



1. Each classifier give probability of instances to be included in each class.
2. Get average of probabilities in each class.
3. $$p(1|x) = \frac{(0.9 + 0.8 + 0.3 + 0.4)}{4}$$

$$p(2|x) = \frac{(0.1 + 0.2 + 0.7 + 0.6)}{4}$$

Stacking

Stacked Generalization

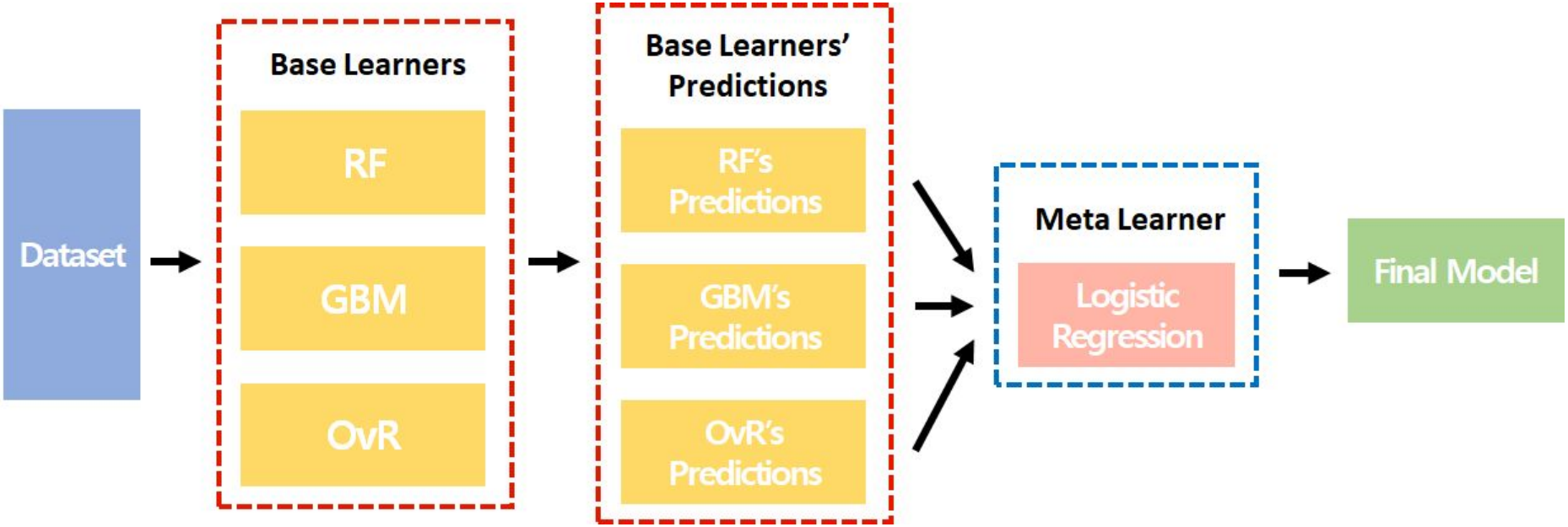
- “Stacked generalization is a generic term referring to any scheme for feeding information from one set of generalizers to another before forming the final guess.” - David H.Wolpert Stacked Generalization”

- Stacking combines base learners by means of a separate meta-learning method using their predictions on held-out data obtained through **cross validation**

- Stacking can be applied to models obtained using different learning algorithms

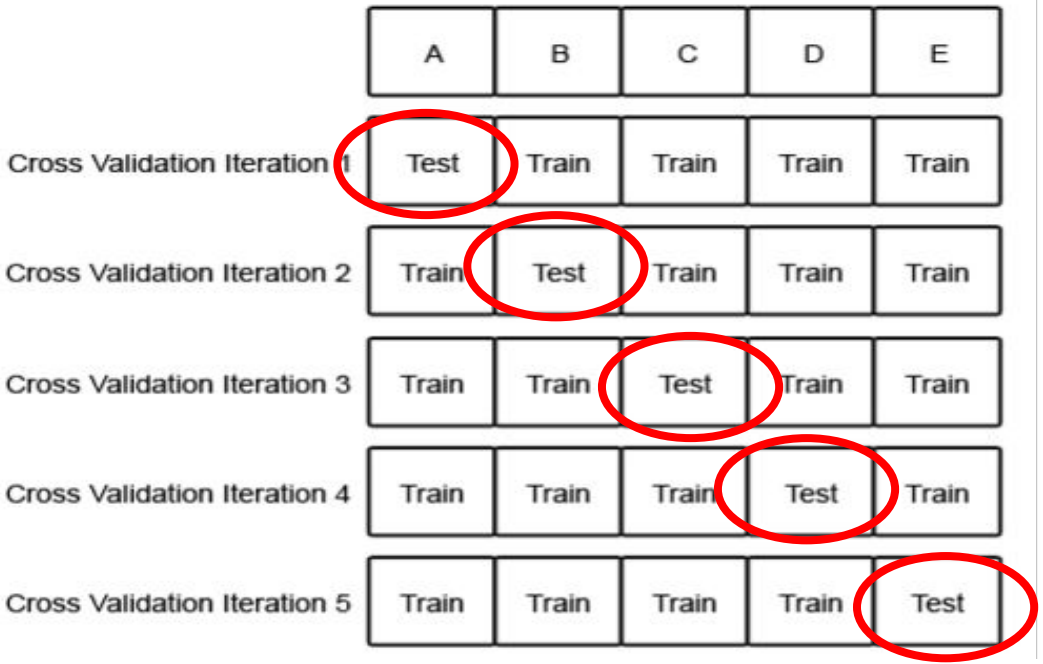
Stacking

General Stacking Process



Cross Validation

K-fold CV

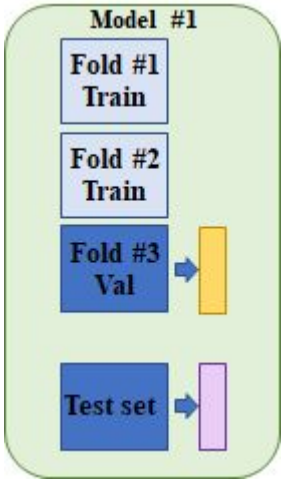


- The original trainset is randomly partitioned into k equal sized subsamples
- Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the other k-1 subsamples are used as training data

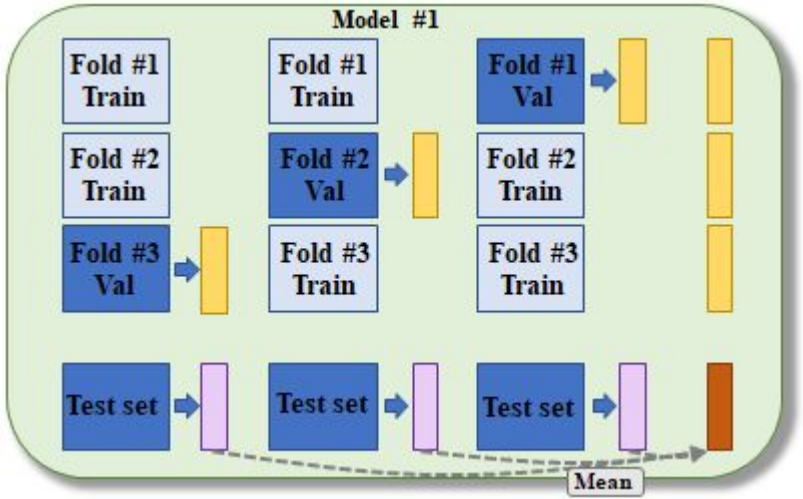
Stacking

CV based Stacking

1.



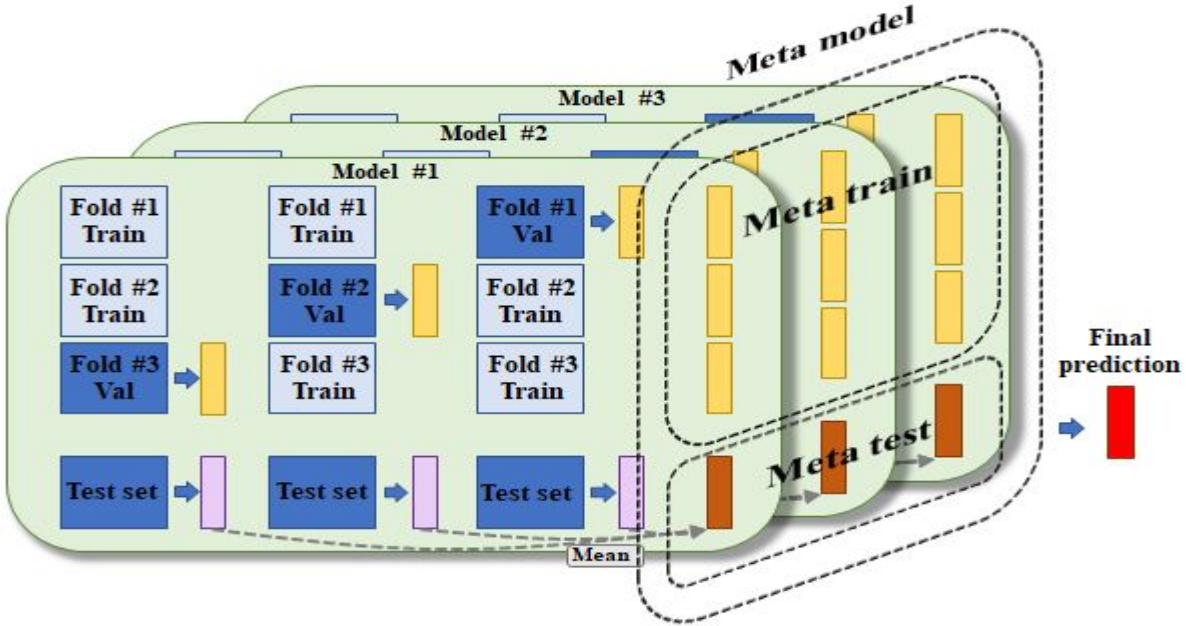
2.



Stacking

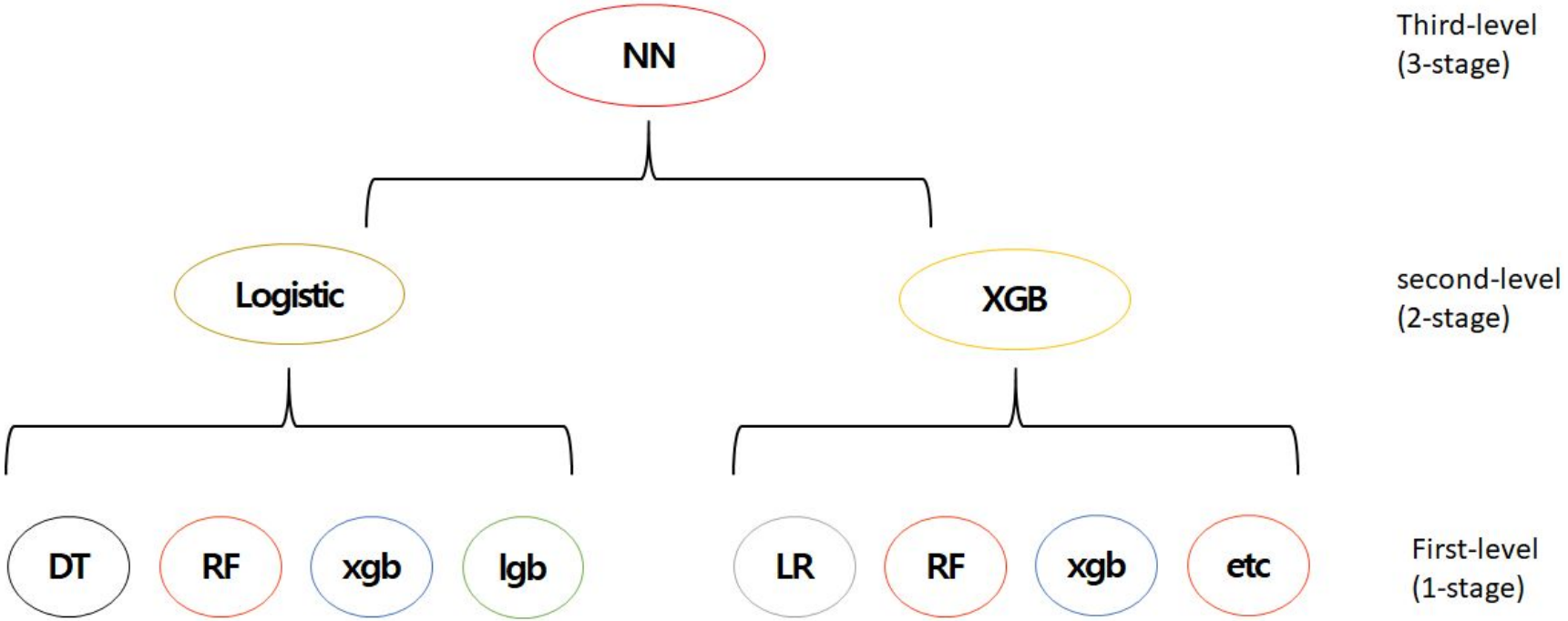
CV based Stacking

3.



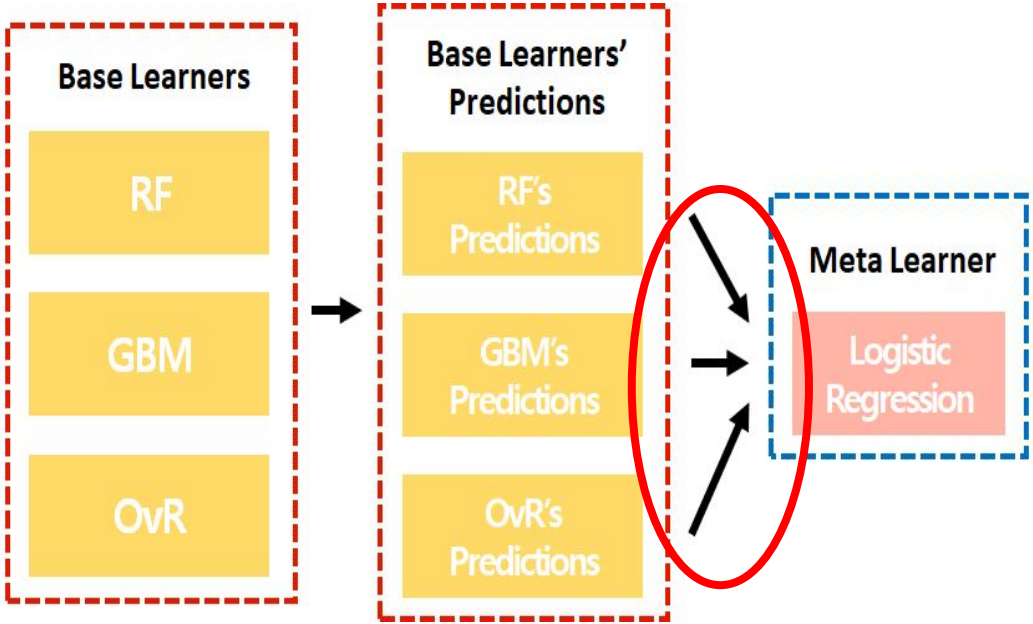
Stacking

Multi-level stacking



Stacking

Stacking Problem



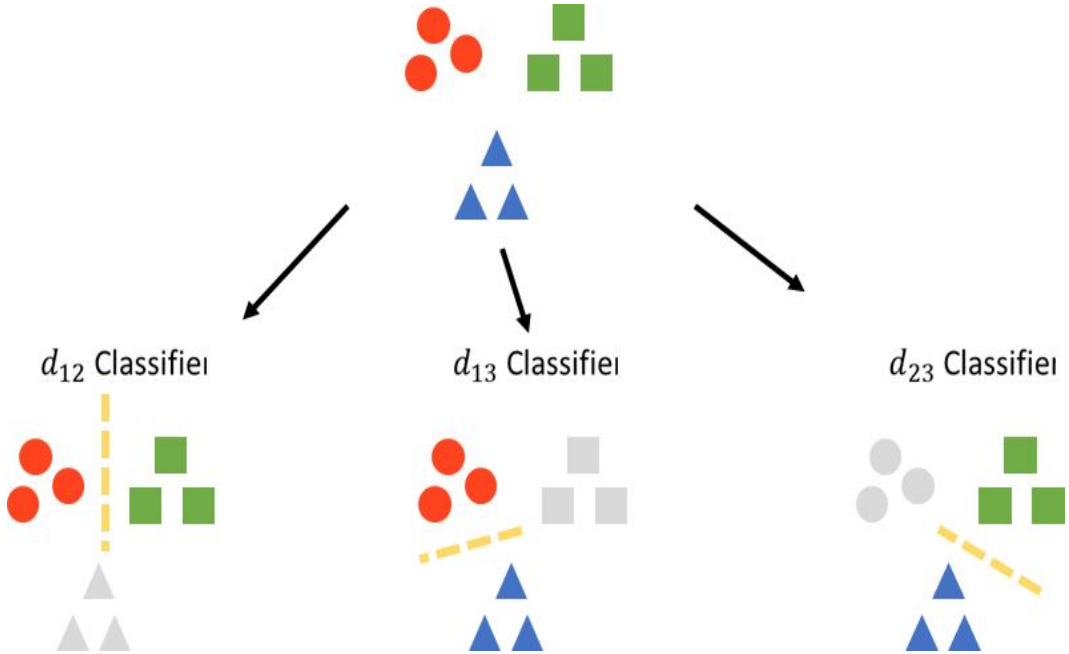
〈How to prevent overfitting problem〉

1. meta learner: linear, logistic

2. when derive the predictions for the meta classifier, do not use the same dataset that was used for training the 1-level classifier
   -〉 Use ONLY the prediction results of the 1-level classifiers

Multi-class

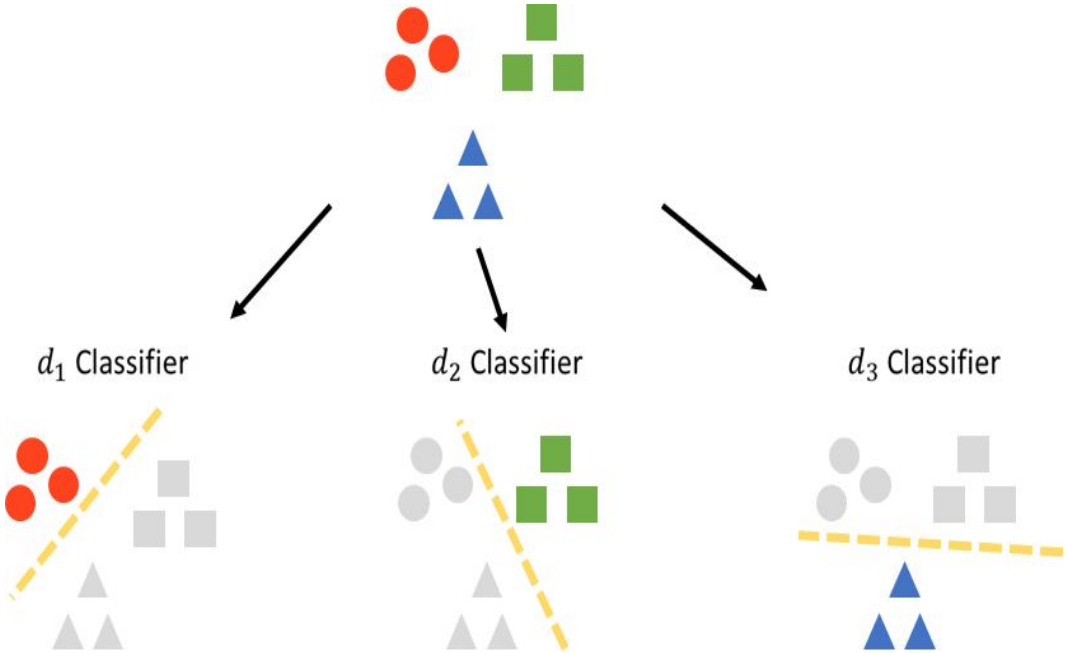
One-vs-One(OvO) Classifier



- If N class exist, train 2 class classifier using samples from the selected two classes only
- $${}_{\mathbf{2}}^{n}\mathrm{C} = \frac{n(n-1)}{2} classifiers$$
- New samples X can be classified by a majority voting process

Multi-class

One-vs-Rest(OvR) Classifier



- Pick one class and train a 2-class classifier with the samples of the selected class on one side and all the other samples on the other side
- N classfiers
- New samples X can be classified by the maximum score amont the N classifiers