

Part.04

Ensemble Learning

# I Stacking

FASTCAMPUS  
ONLINE

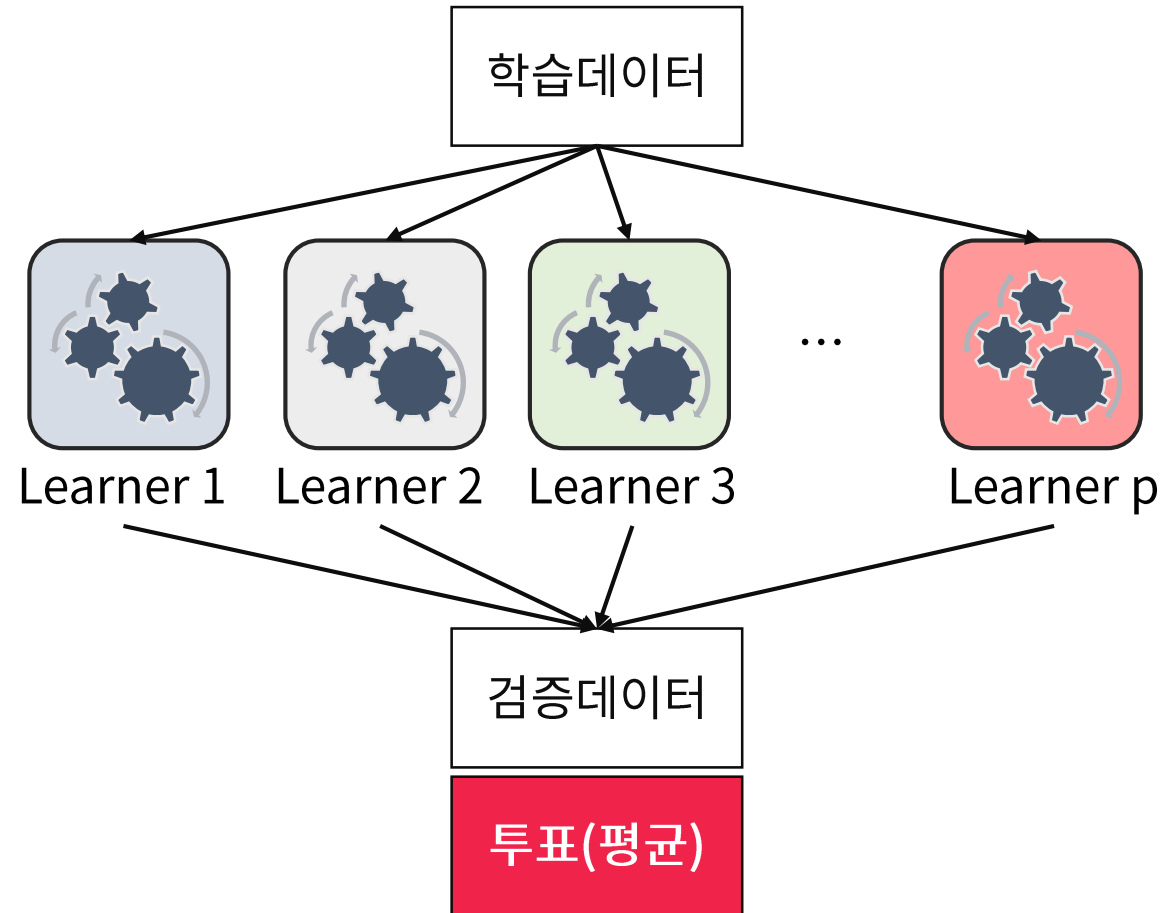
머신러닝과 데이터분석 A-Z

강사. 이경택

# I Stacking

## ■ Ensemble Learning

- 여러 개의 기본 모델을 활용하여 하나의 새로운 모델을 만들어내는 개념



# I Stacking

## ■ Ensemble Learning의 종류

- Bagging : 모델을 다양하게 만들기 위해 데이터를 재구성
- RandomForest : 모델을 다양하게 만들기 위해 데이터 뿐만 아니라, 변수도 재구성
- Boosting : 맞추기 어려운 데이터에 대해 좀더 가중치를 두어 학습하는 개념  
Adaboost, Gradient boosting (Xgboost, LightGBM, Catboost)

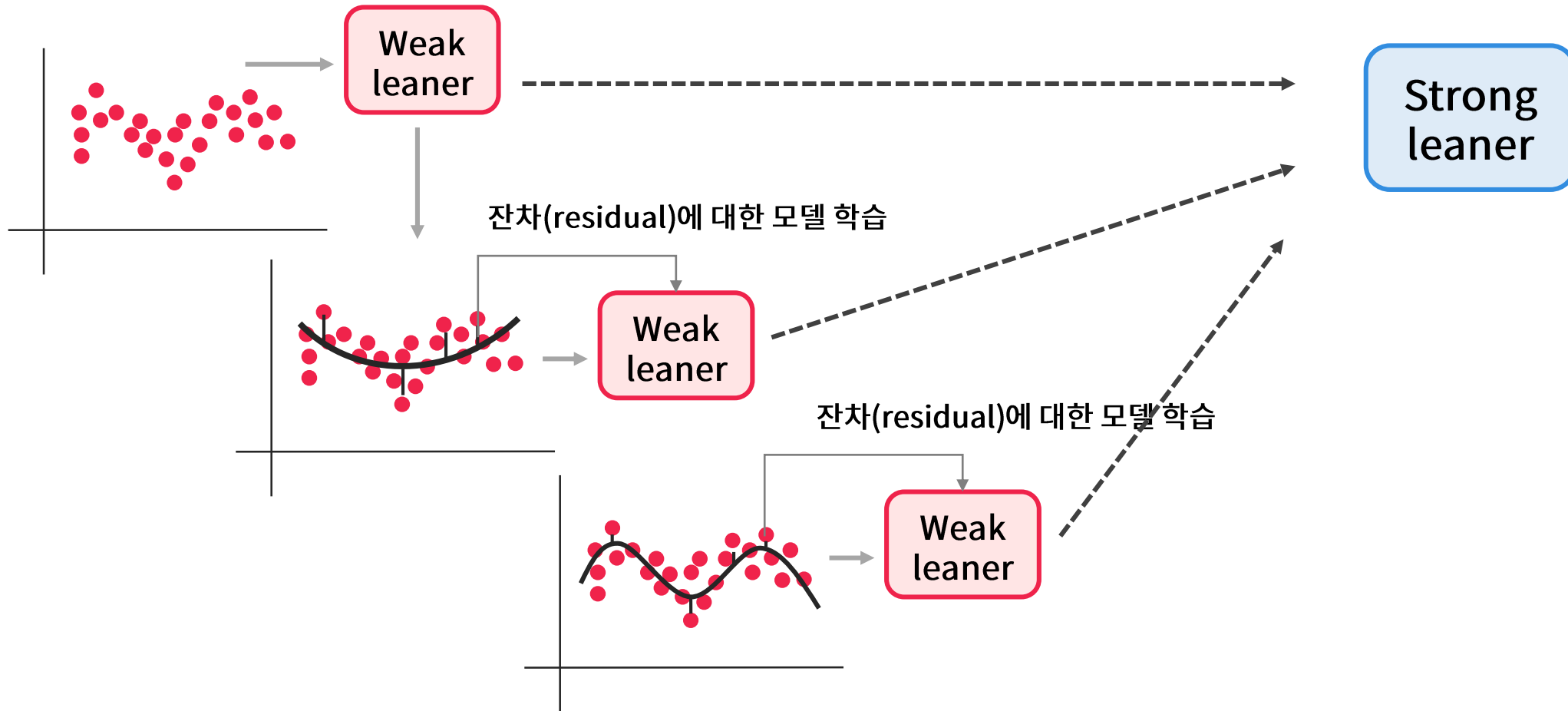
Tree기반의 단일 모델  
(패키지 함수)

- Stacking : 모델의 output값을 새로운 독립변수로 사용

“Ensemble의 한 개념”

# I Stacking

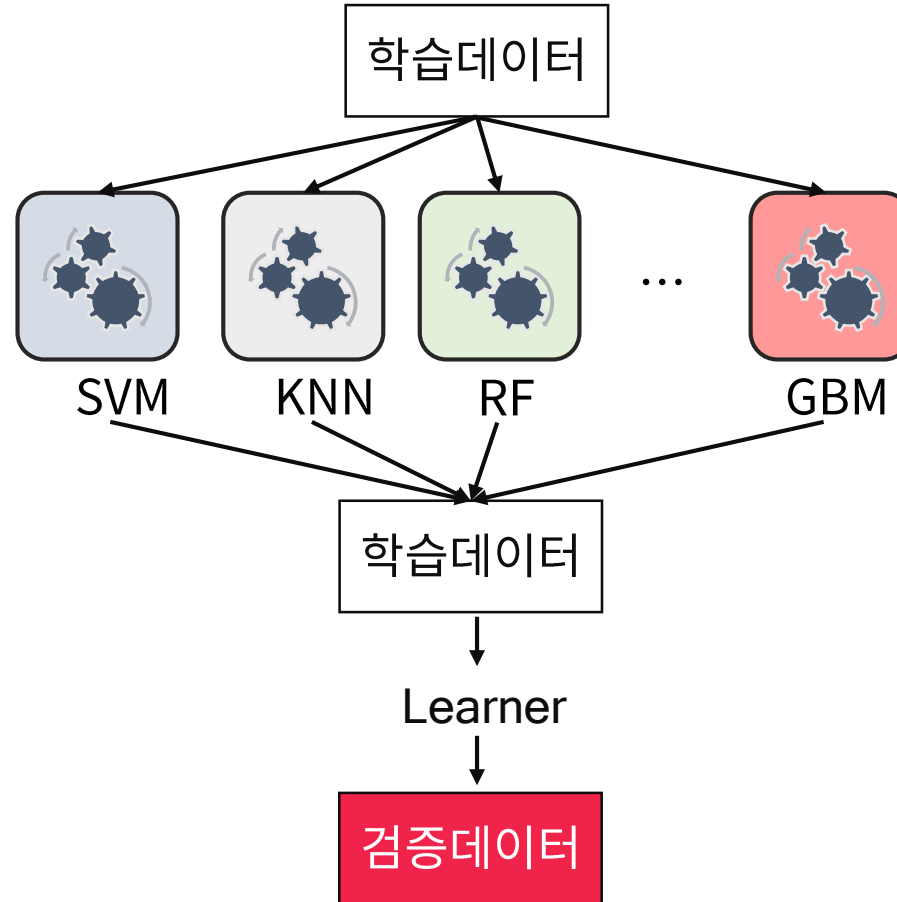
## ■ Gradient Boosting



# I Stacking

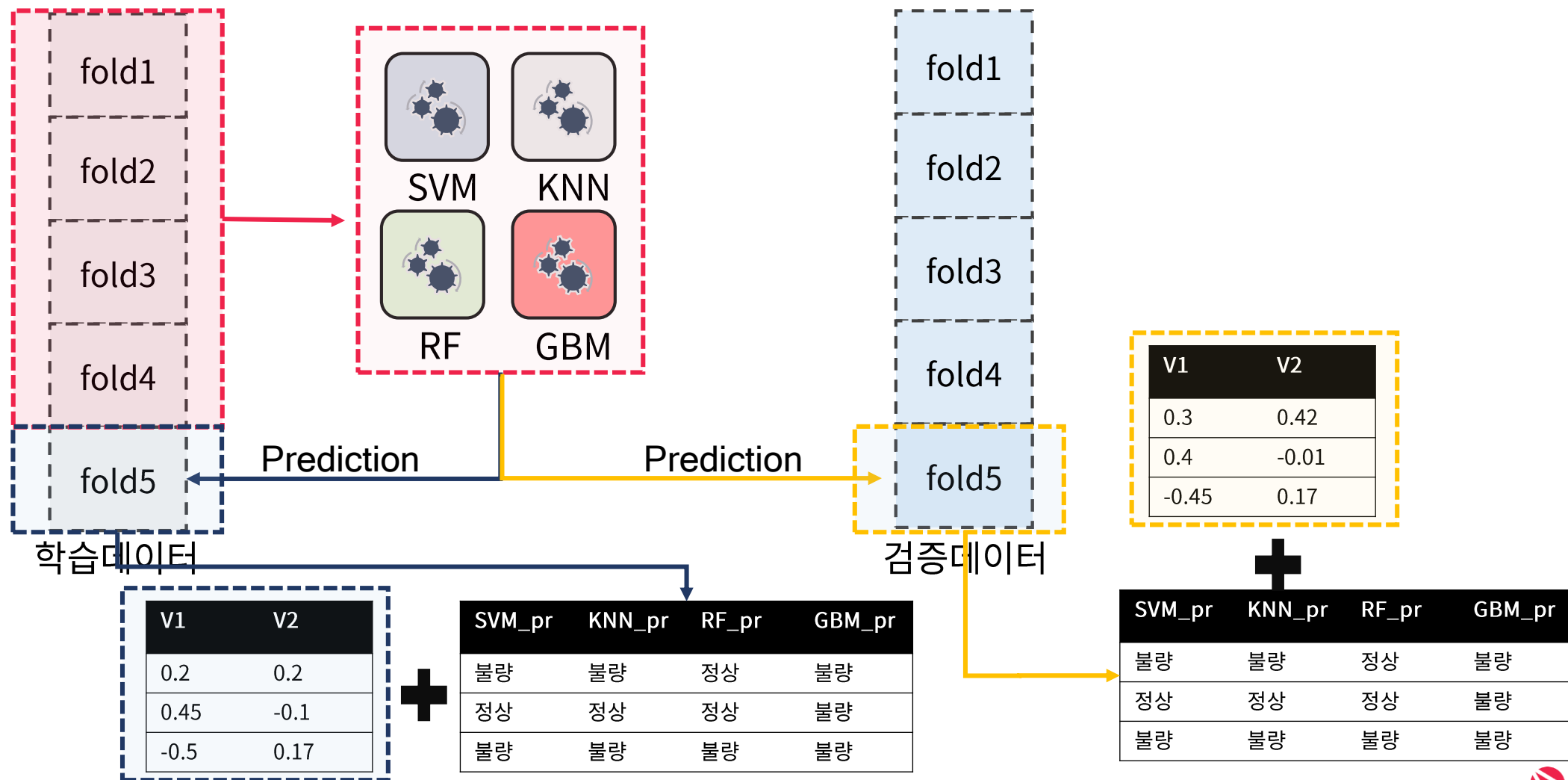
## ■ Stacking이란

- Meta Learner라고 불리우며, 다양한 모델을 결합하여 사용하는 기법



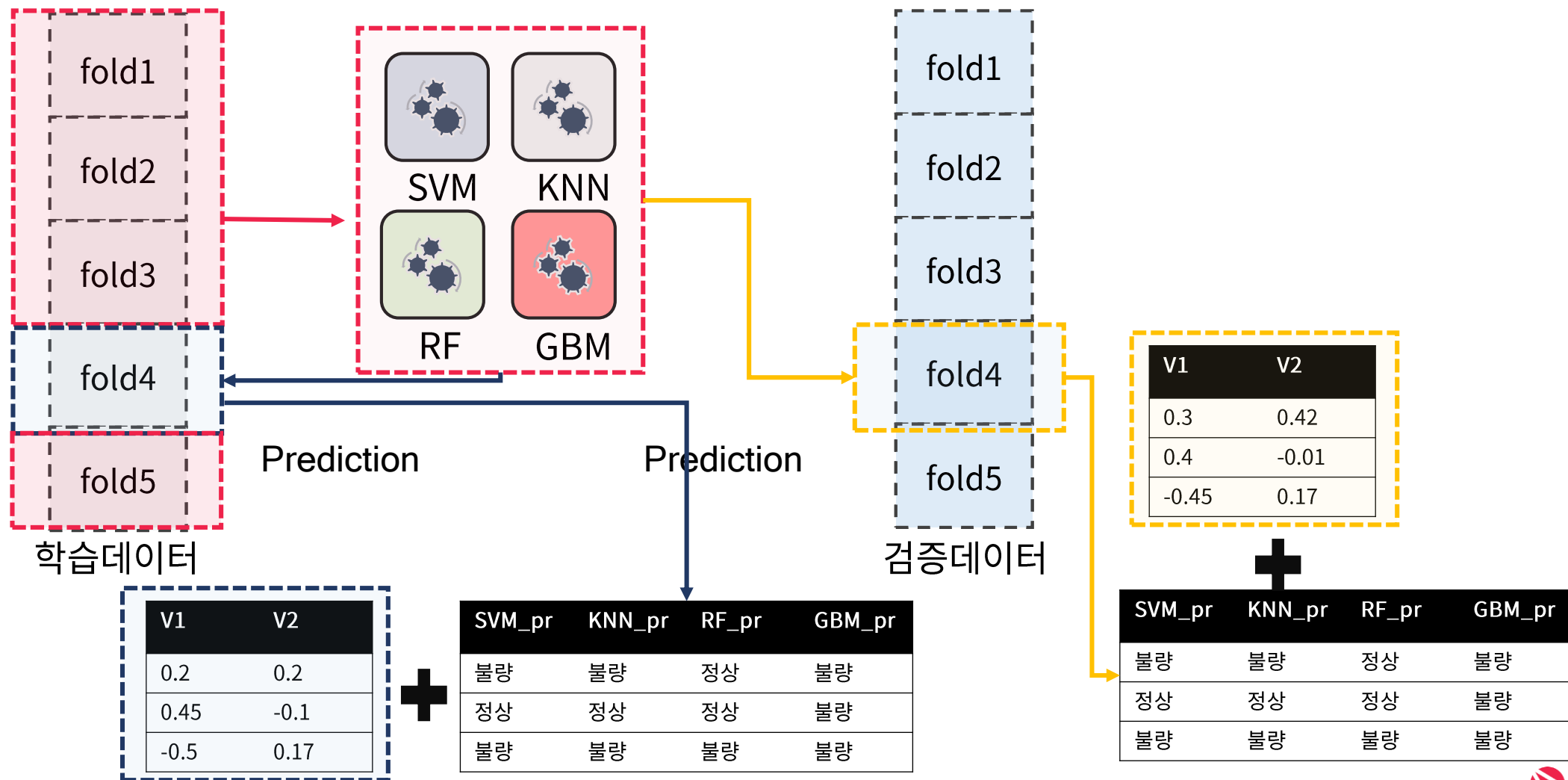
# I Stacking

## Stacking이란



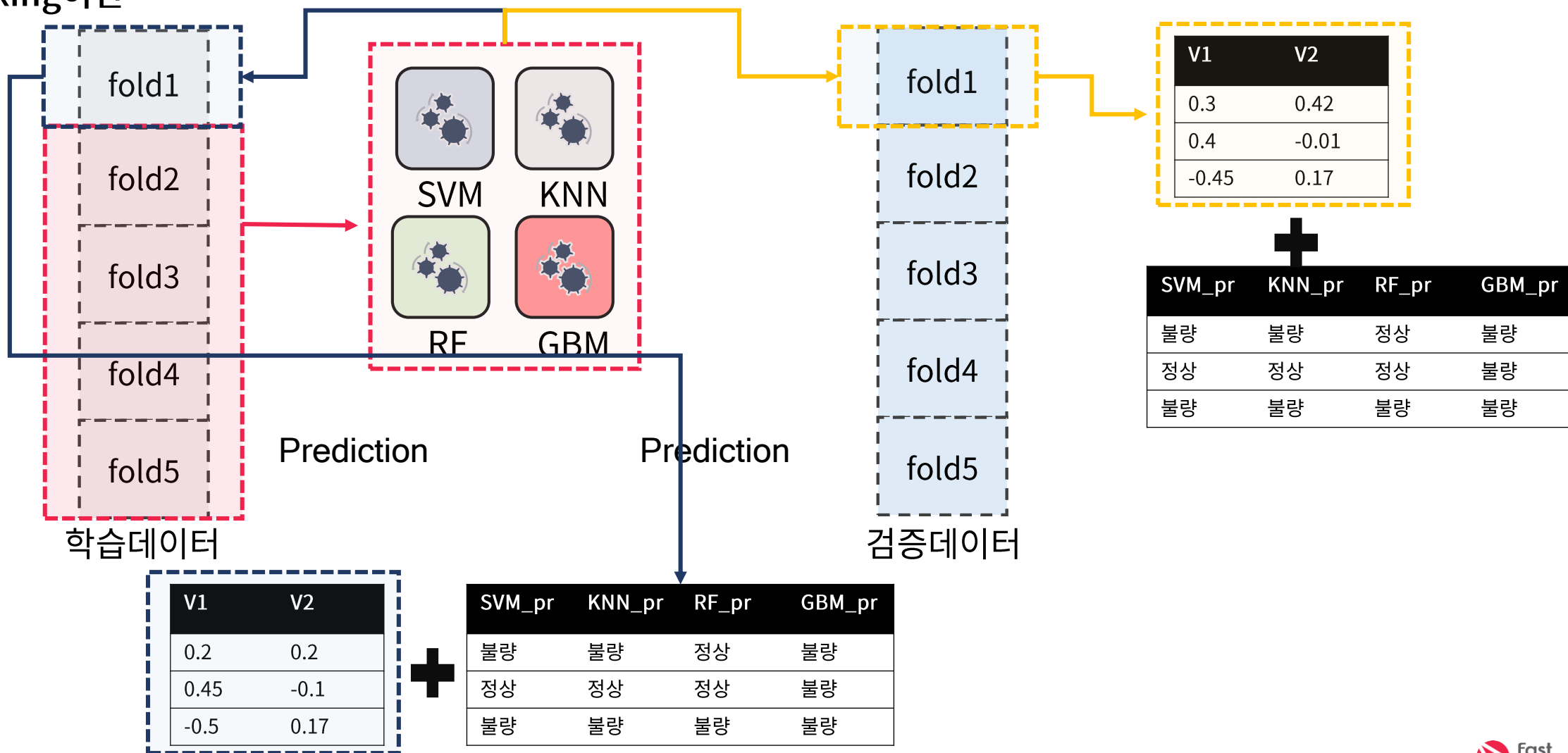
# I Stacking

## Stacking이란



# I Stacking

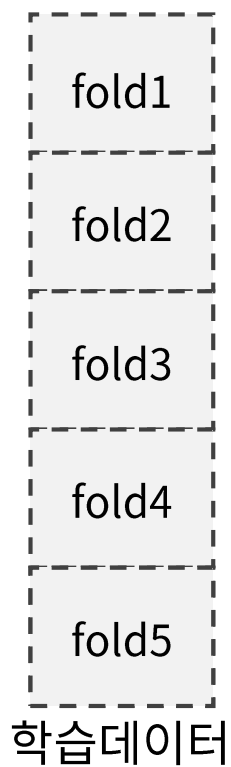
## Stacking이란





# I Stacking

## Stacking이란



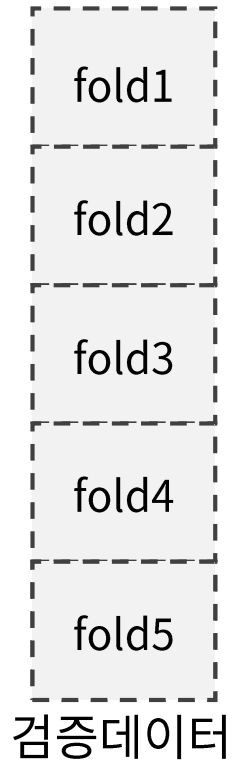
새로운 학습데이터

V1	V2		SVM_pr	KNN_pr	RF_pr	GBM_pr
0.2	0.2	+	불량	불량	정상	불량
0.45	-0.1		정상	정상	정상	불량
0.2	0.2		불량	불량	정상	불량
0.45	-0.1		정상	정상	정상	불량
0.2	0.2		불량	불량	정상	불량
0.45	-0.1		정상	정상	정상	불량
0.2	0.2		불량	불량	정상	불량
0.45	-0.1		정상	정상	정상	불량
0.2	0.2		불량	불량	정상	불량
0.45	-0.1		정상	정상	정상	불량
-0.5	0.17		불량	불량	불량	불량

기존 학습데이터      각 모델별 prediction 값

# I Stacking

## ■ Stacking이란



새로운 검증데이터

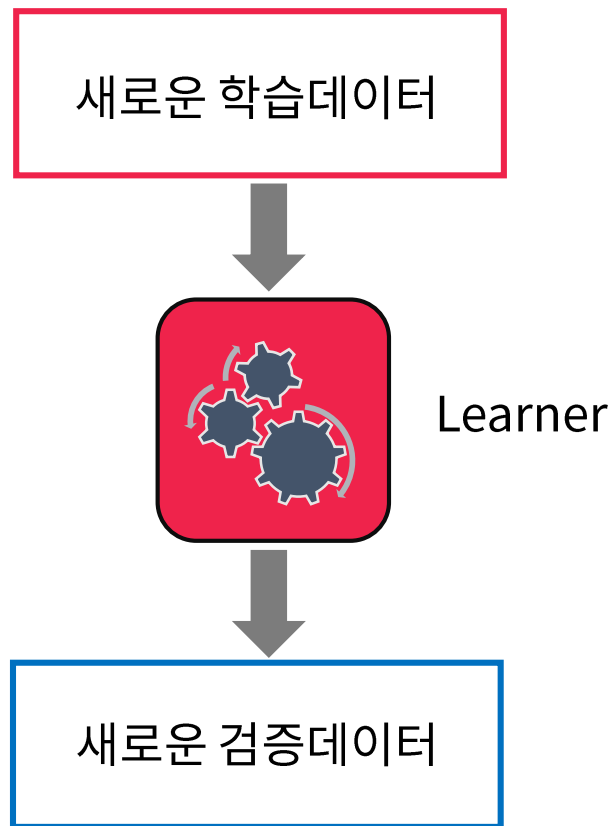
V1	V2	+	SVM_pr	KNN_pr	RF_pr	GBM_pr
0.2	0.2		불량	불량	정상	불량
0.4	-0.1		정상	정상	정상	불량
0.12	0.12		불량	불량	정상	불량
0.45	-0.1		정상	정상	불량	정상
0.2	0.12		불량	정상	정상	불량
0.35	-0.15		정상	정상	정상	정상
0.21	0.21		불량	불량	정상	불량
0.45	-0.1		정상	불량	불량	불량
0.02	0.2		불량	불량	정상	불량
0.45	-0.1		정상	정상	정상	정상
-0.5	0.17		불량	불량	불량	불량

기존검증데이터

각 모델별 prediction 값

# I Stacking

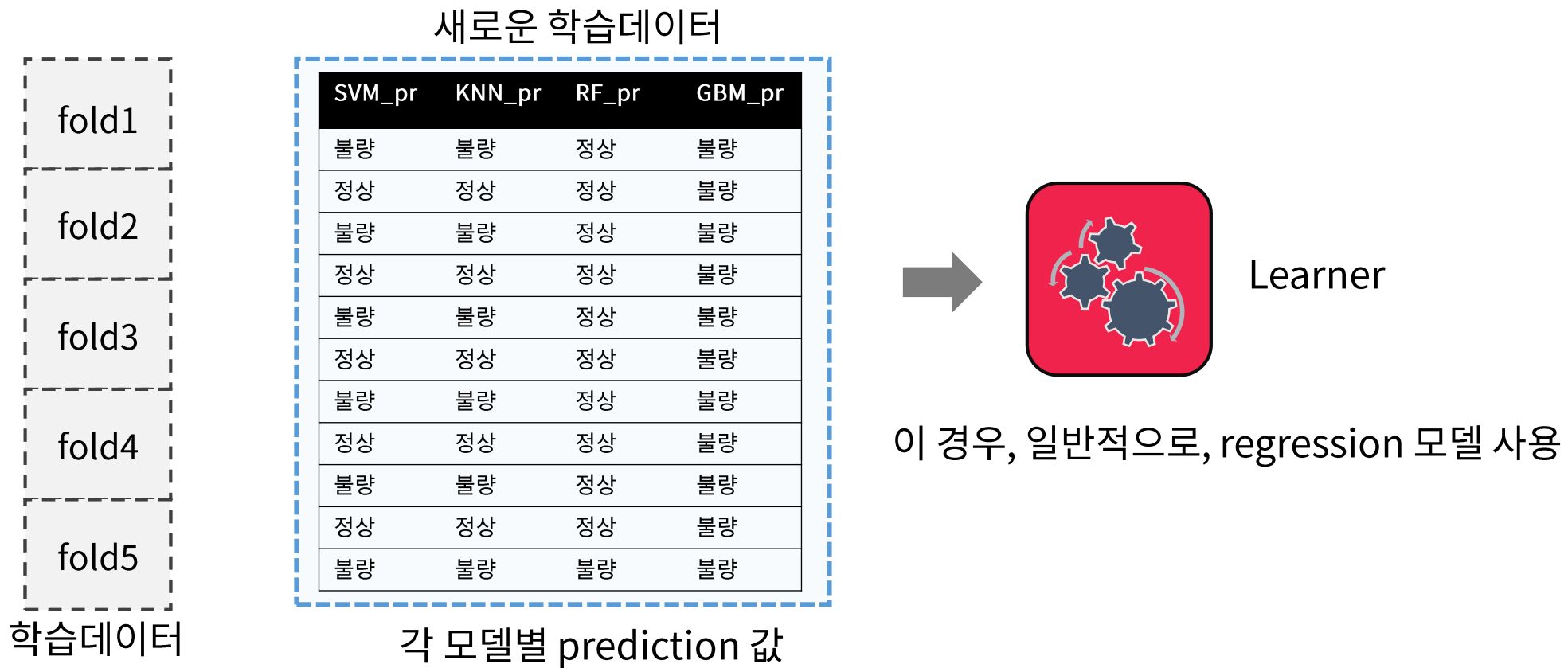
- Stacking이란



# I Stacking

## ■ Stacking이란

- 기존 feature를 쓰지 않고 각 모델별 prediction만을 사용하기도함



Part.04

Ensemble Learning

# I Ensemble의 Ensemble

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택