

Tutorial

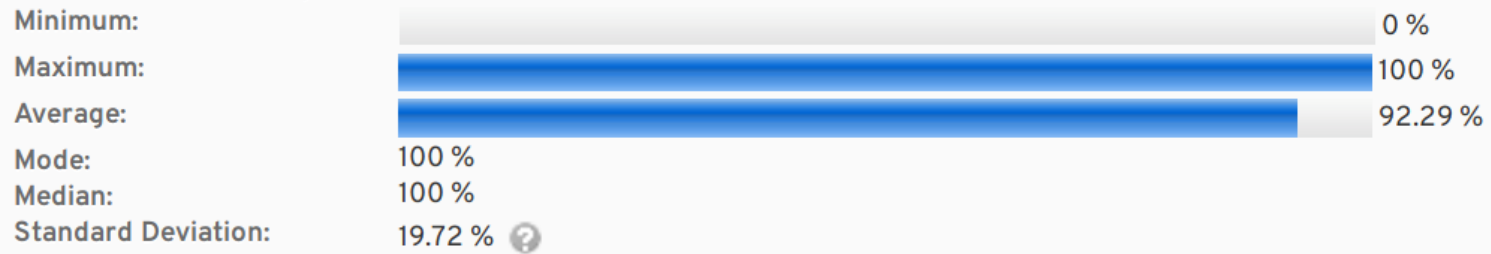
Cross Validation, Bagging, Ensemble

- By Shivam

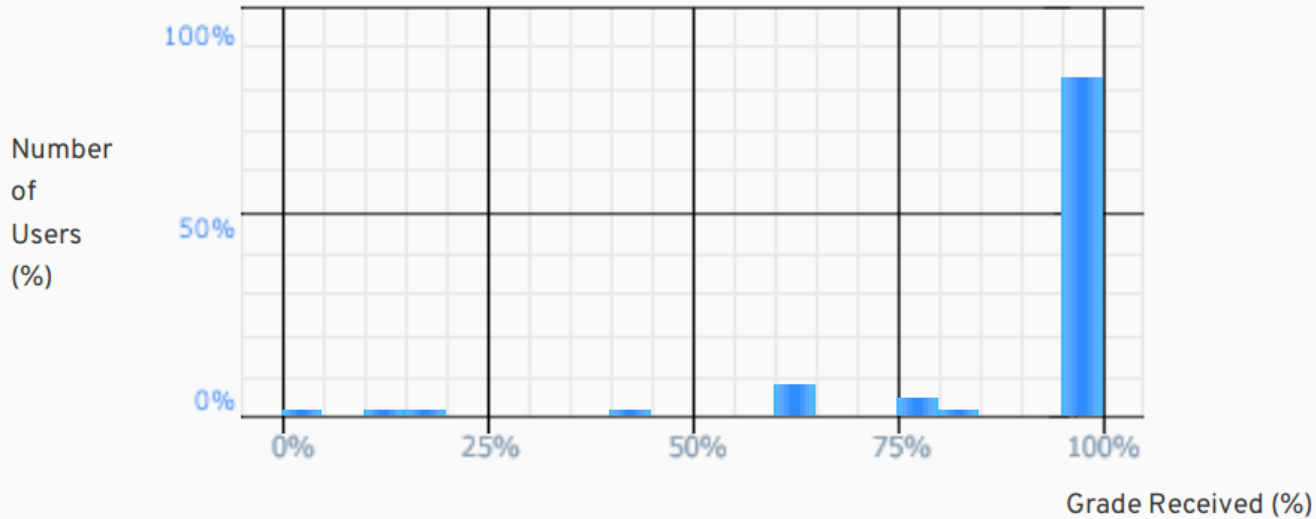
Assignment 2

Tutorial Assignment 2 Class Statistics

Number of submitted grades: 96 / 106



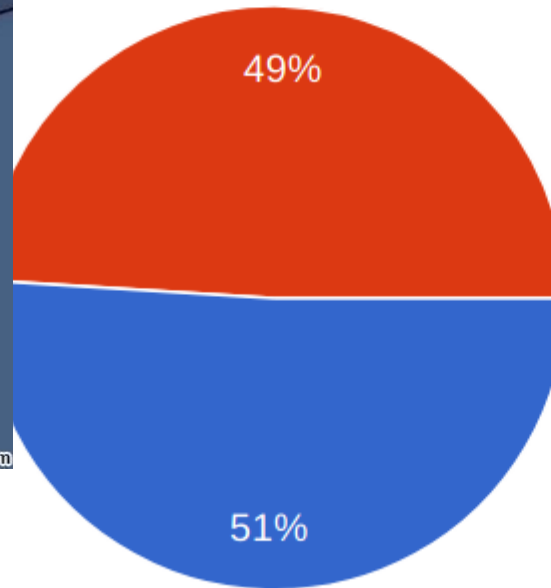
Grade Distribution



Assignment 3 - Survey

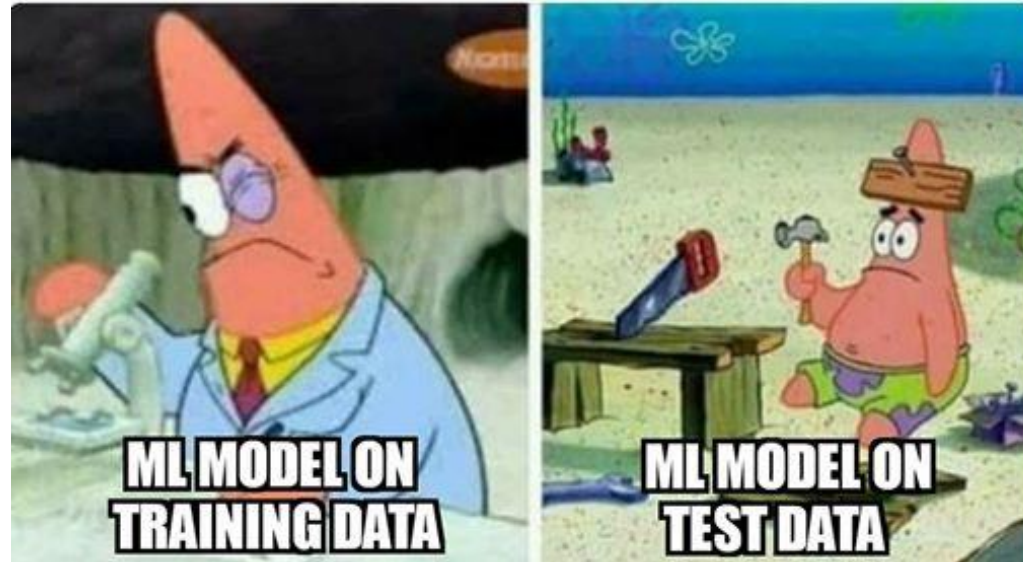
How you want to do assignment 3?

51 responses



- Individually
- In a group (max 2)

ML in summary...



Different scenarios...

- Training Accuracy: 97%
 - Test Accuracy: 45%
-
- Training Accuracy: 97%
 - Test Accuracy: 92%

Training ML models

- Are they generalizing?
- Are you overfitting?
- “Babysit” training of your ML models
- Perform hyper parameter tuning

They sound trivial but very vital for training AWESOME models

You don't want this..



Today's tutorial

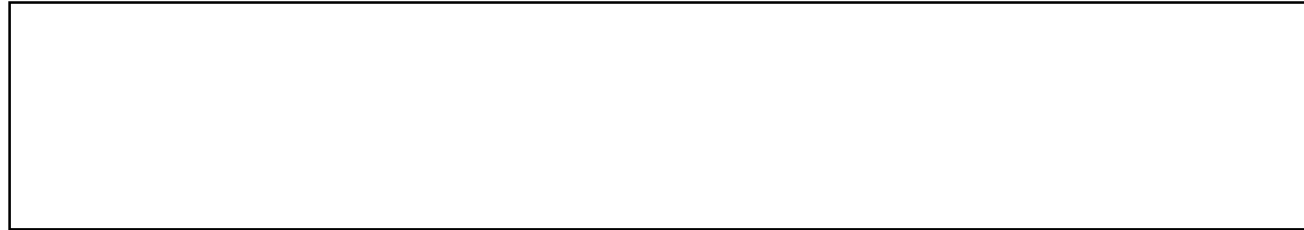
- Will talk techniques to avoid overfitting
- Different validation techniques during training
- Bagging to prevent over fitting
- Ensemble for better generalization
- Python implementation for all of them

Very important techniques to score good on your next assignment!!

Cross Validation

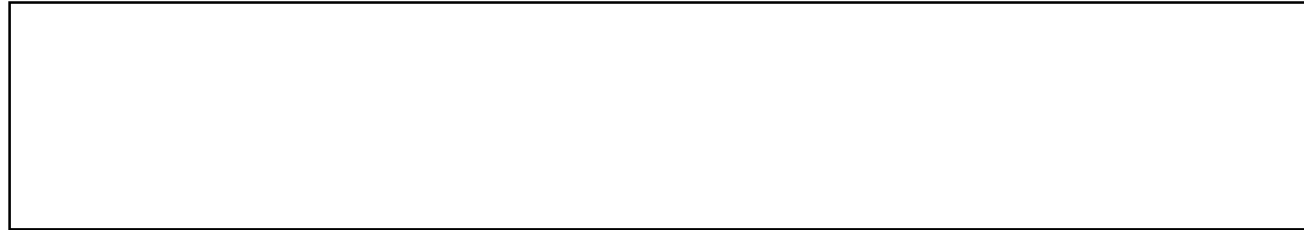
- Estimate accuracy while training
- Early stopping
- Hyper parameter tuning

K-Fold Cross Validation

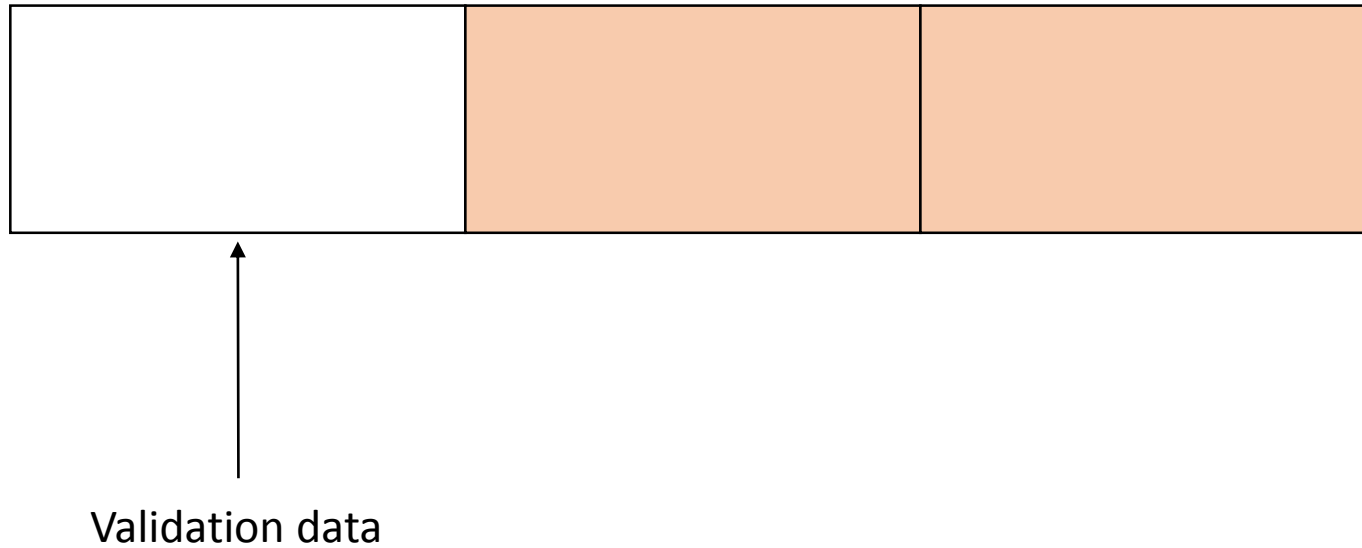


K-Fold Cross Validation

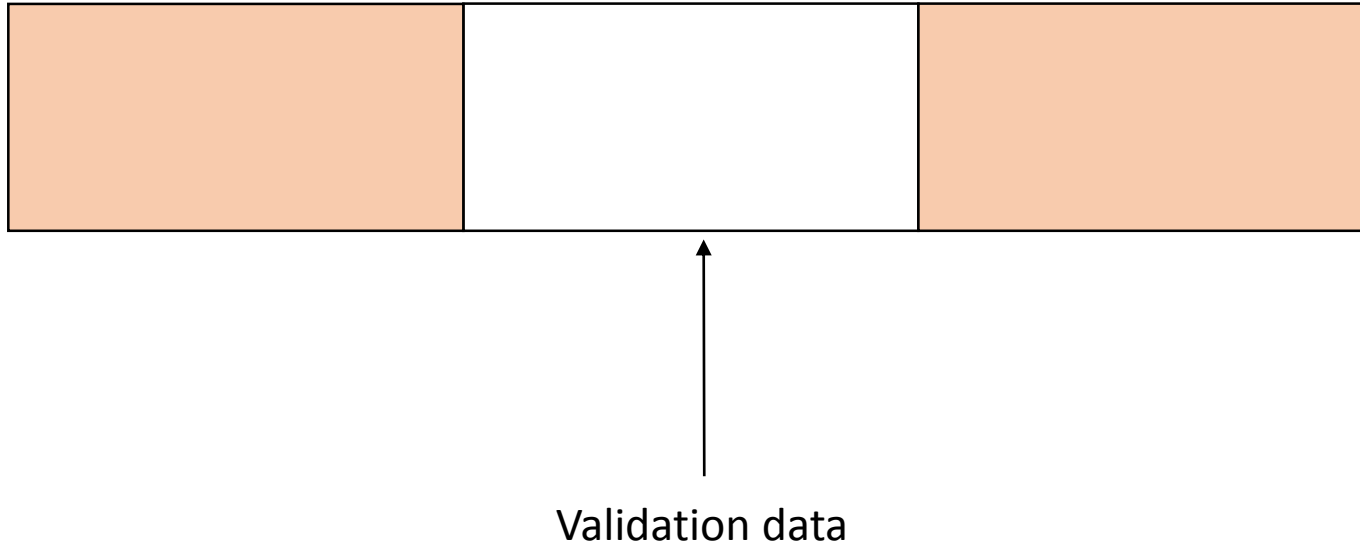
Random Shuffle



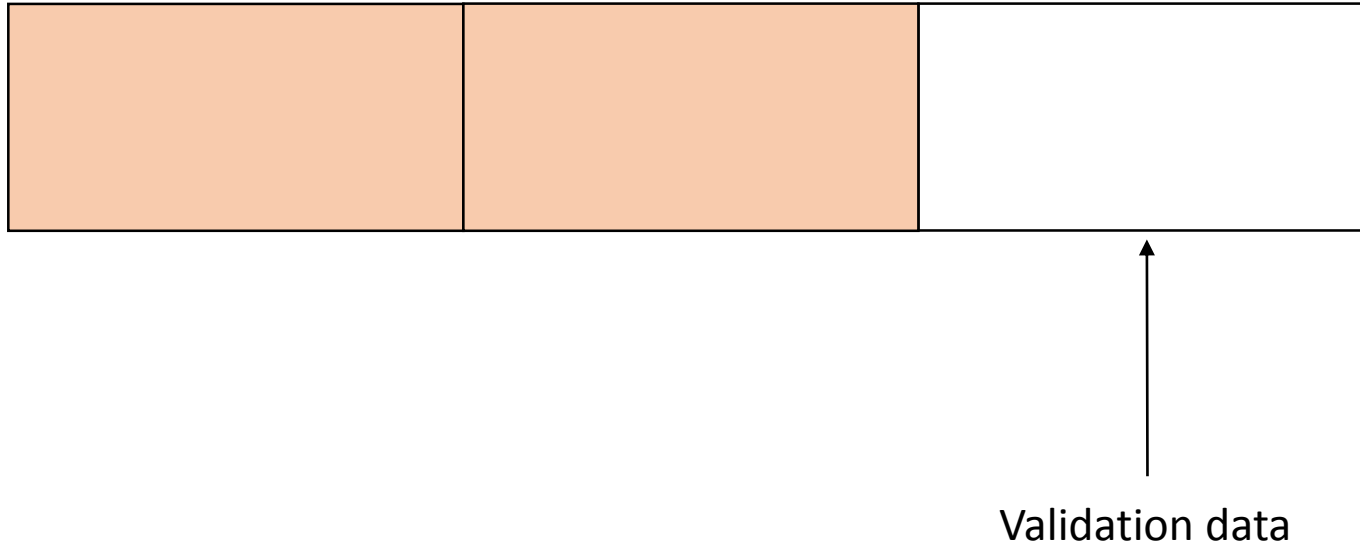
K-Fold Cross Validation



K-Fold Cross Validation



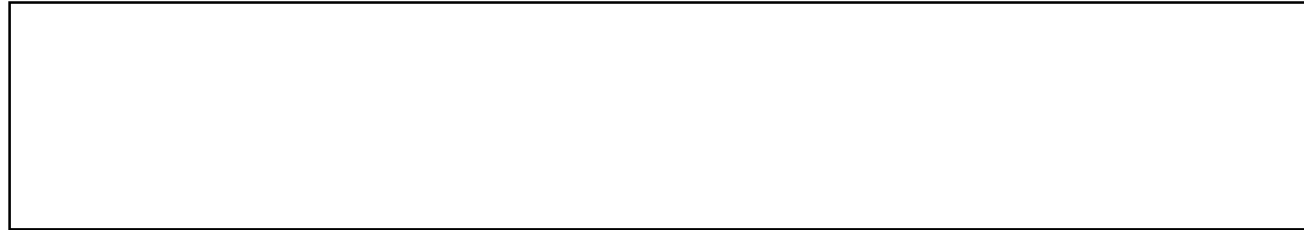
K-Fold Cross Validation



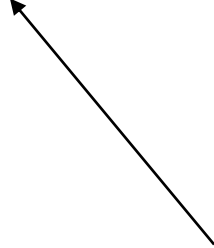
K-Fold Cross Validation

- Generalize better
- Better than fixed validation set

Leave one out



Leave one out



One "thing" in dataset

Leave one out



One “thing” in dataset

Leave one out



One “thing” in dataset

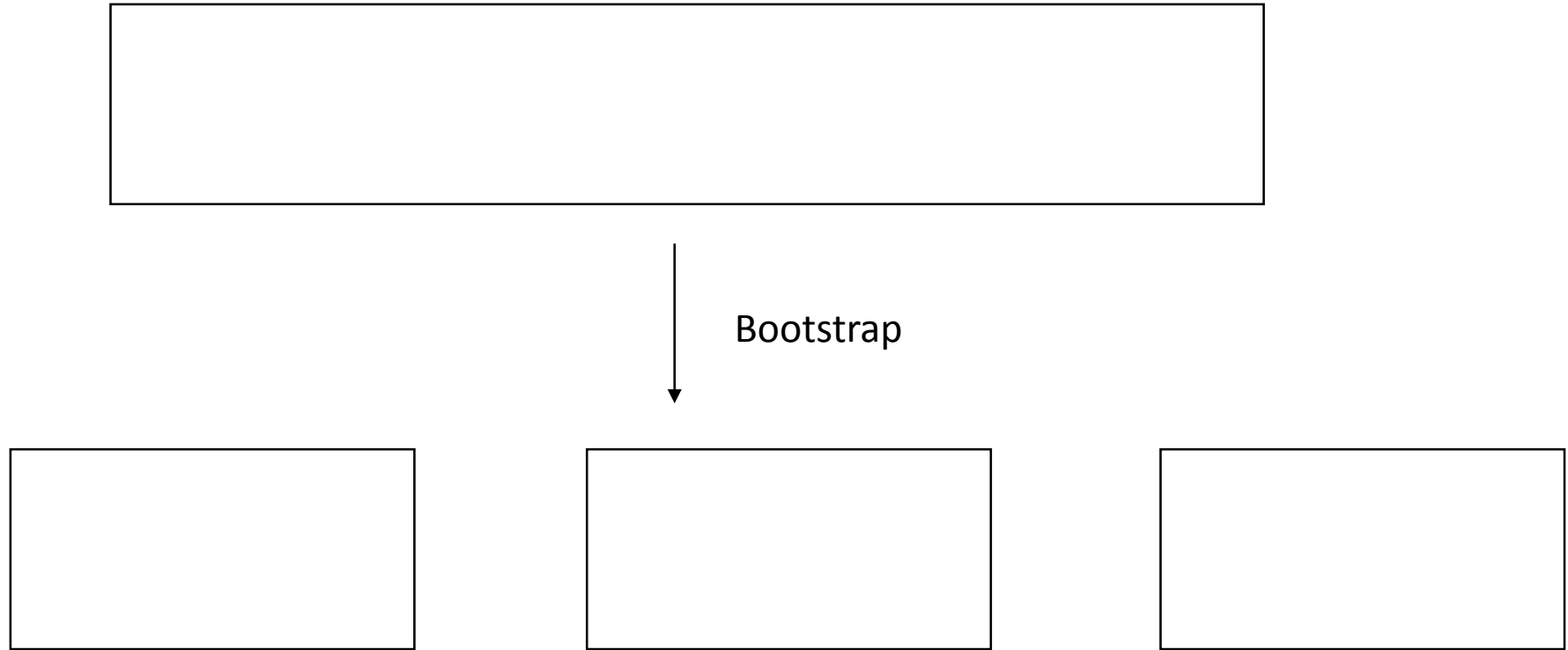
LOO vs KF

- Is KF generalized LOO?
- How?

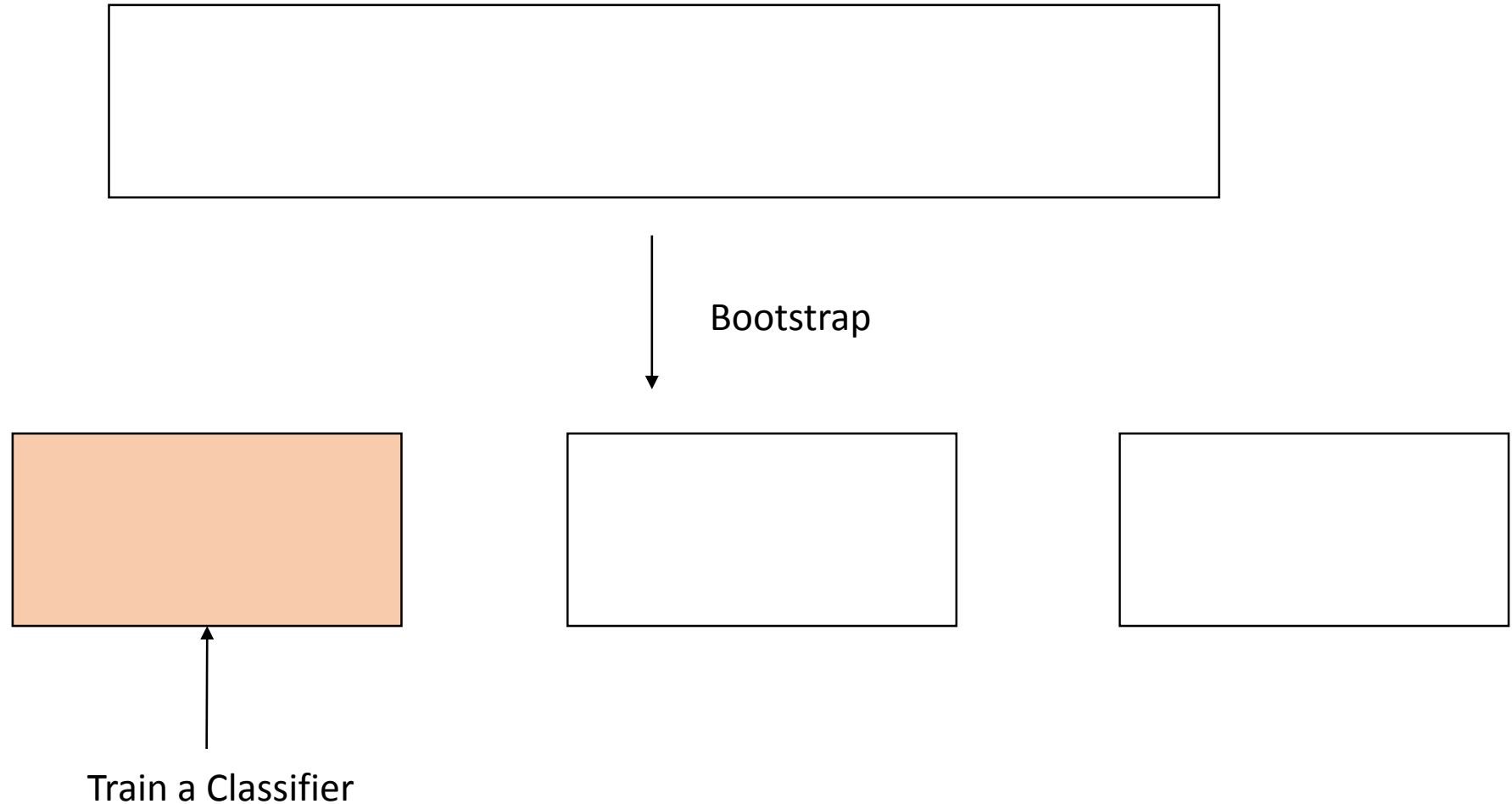
Bagging: Bootstrap Aggregation



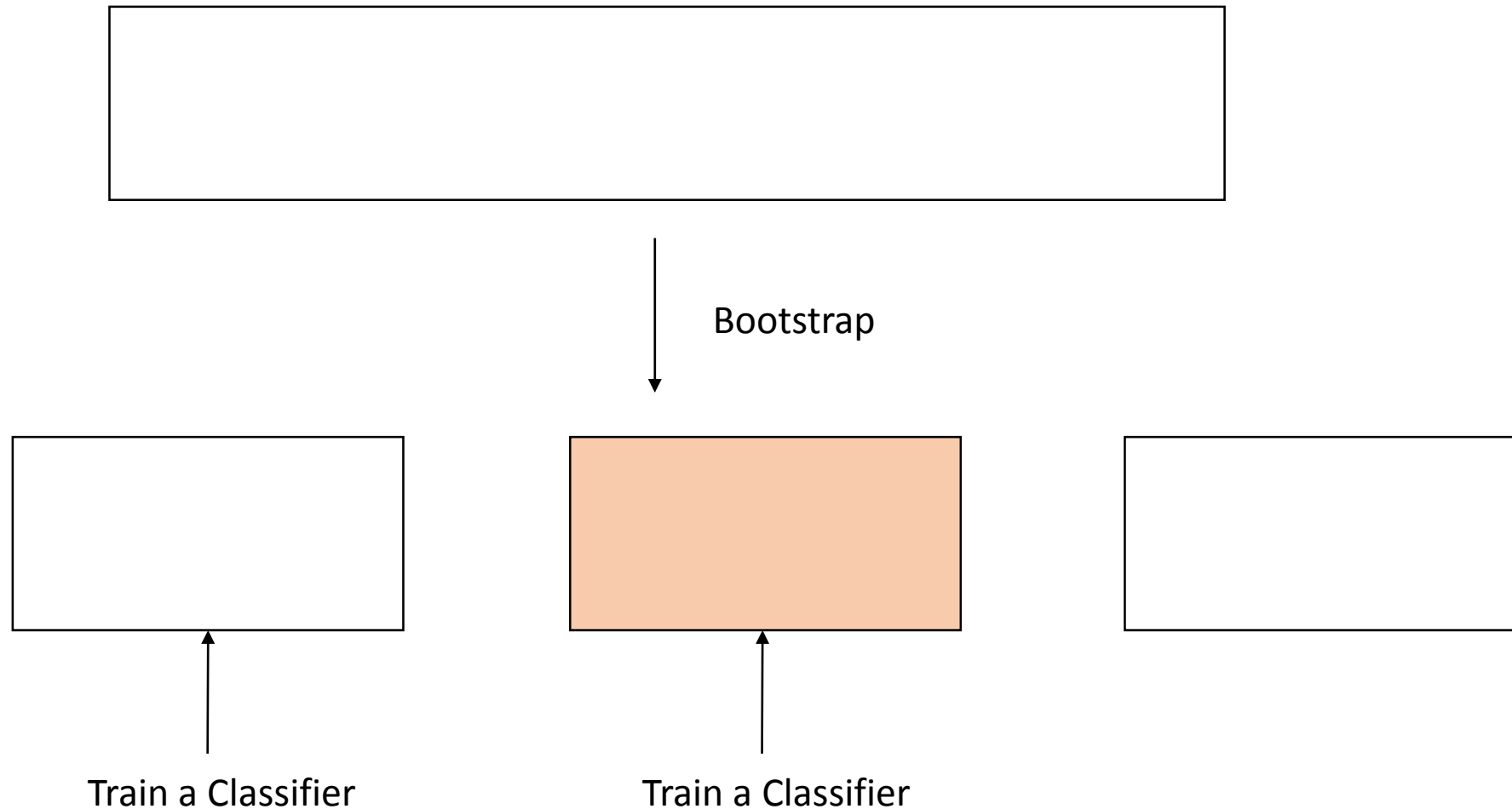
Bagging: Bootstrap Aggregation



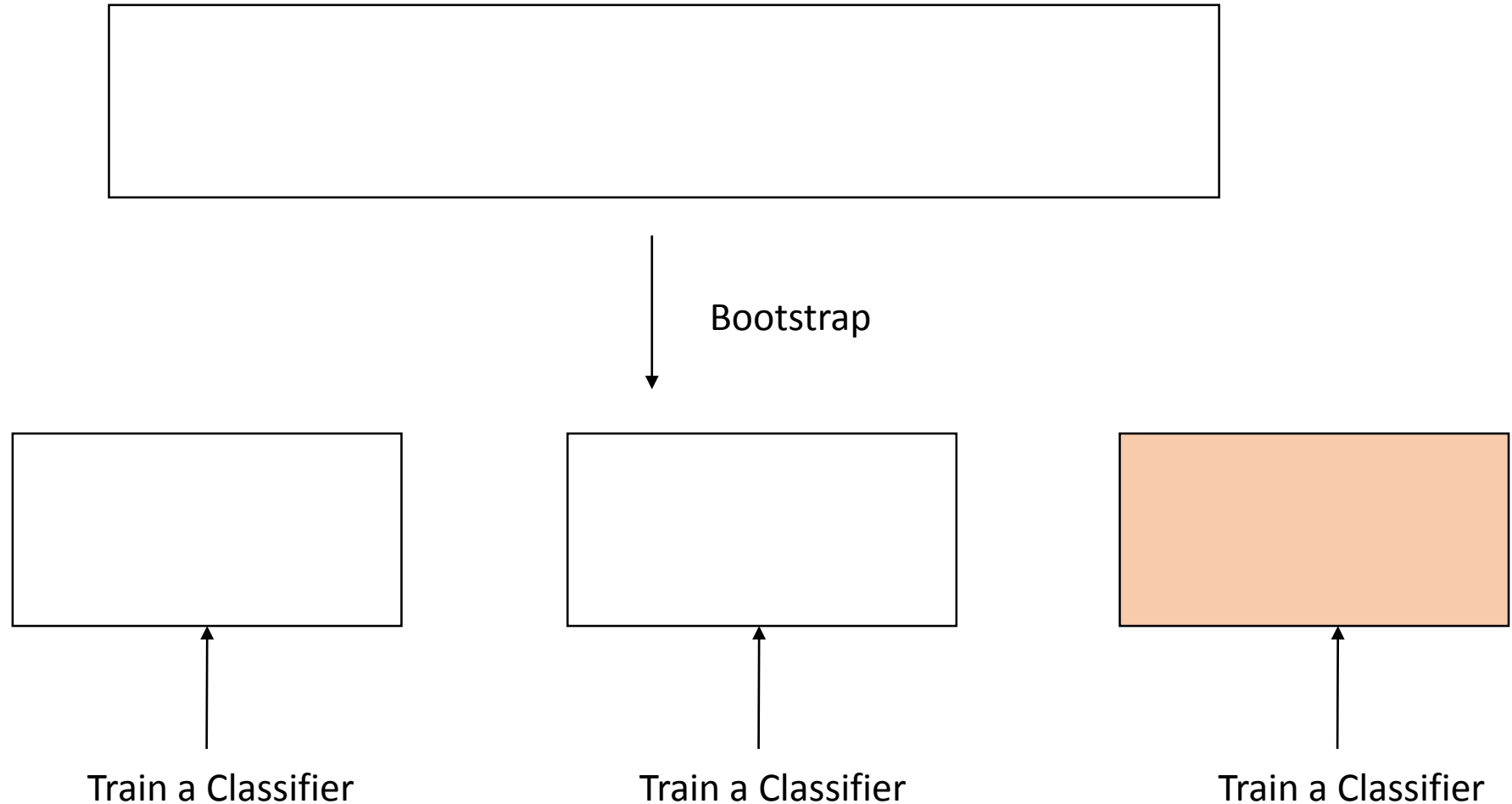
Bagging: Bootstrap Aggregation



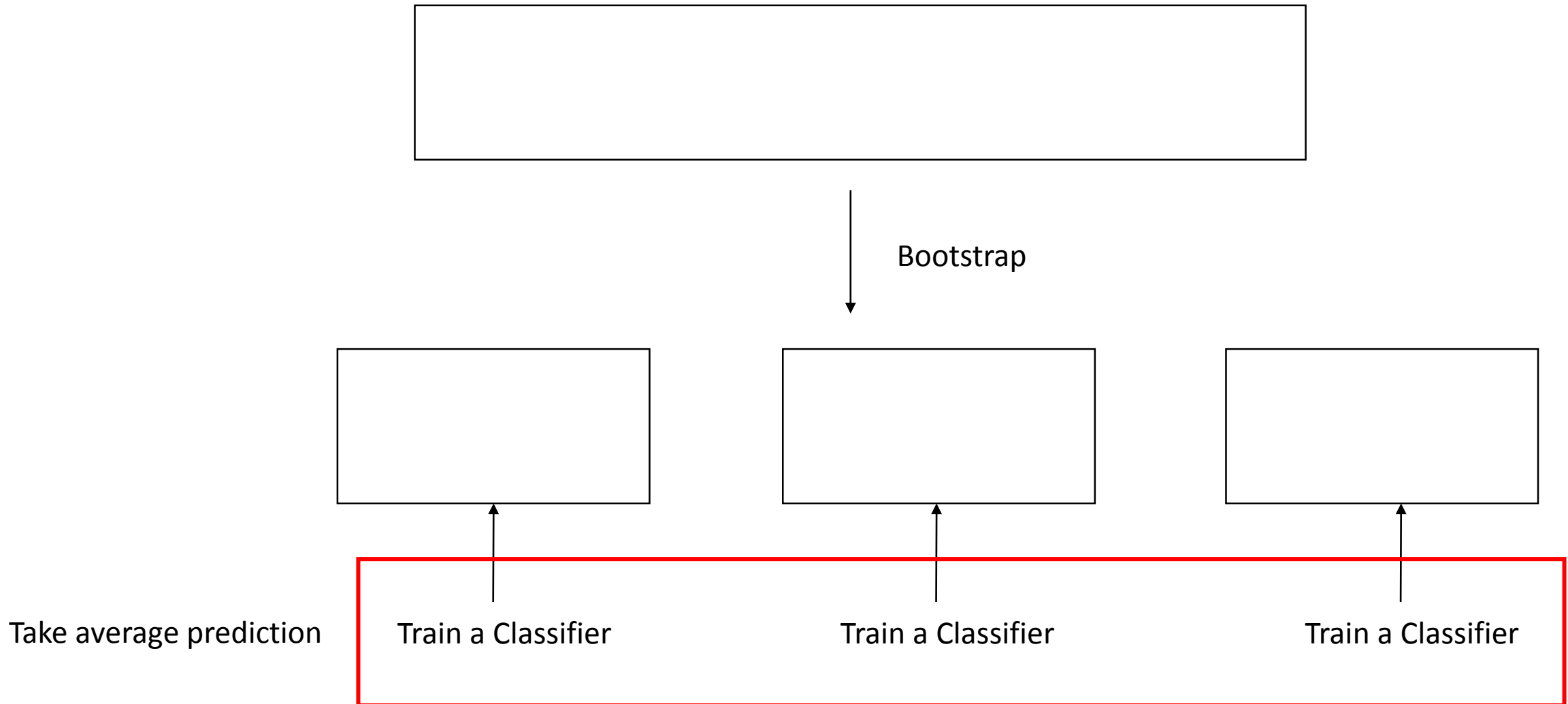
Bagging: Bootstrap Aggregation



Bagging: Bootstrap Aggregation



Bagging: Bootstrap Aggregation



Why does it even work?

- Anyone take a guess?

Why does it even work?

God of randomness



“Expose yourself to as much randomness as possible.”

Ensemble

- It just works
- Personal favorite to score better on data competitions
- Kind of “trick” to score better
- Again it works due to “randomness”

Ensemble – in essence

- Train bunch of different classifiers
- Take average during predictions
- You can train classifier using bagging or k-fold or as you may like