Stacking Method for Classification

Tatsuhiro Eguchi, Affiliation
2IE23028R

PR

PR

Purpose of Exercise

- Dataset
  - Titanic Dataset
- Reason
  - Simple Table Data
  - Famous for Kaggle competition
  - 25th Anniversary Film Release Celebration
- Purpose
  - Achieve high accuracy
  - Compare Multiple Classification Methods

PR

Approach

- Classification methods
  - K-nearest neighbors(KNN)
  - Extra Tree(EXT)
  - Random forest(RFC)
  - Gradient Boosting(GBC)
  - Extreme Gradient Boosting(XGB)

  - **Ensemble Learning model(Original)**
    - Stacking method
    - Combine the above classification methods

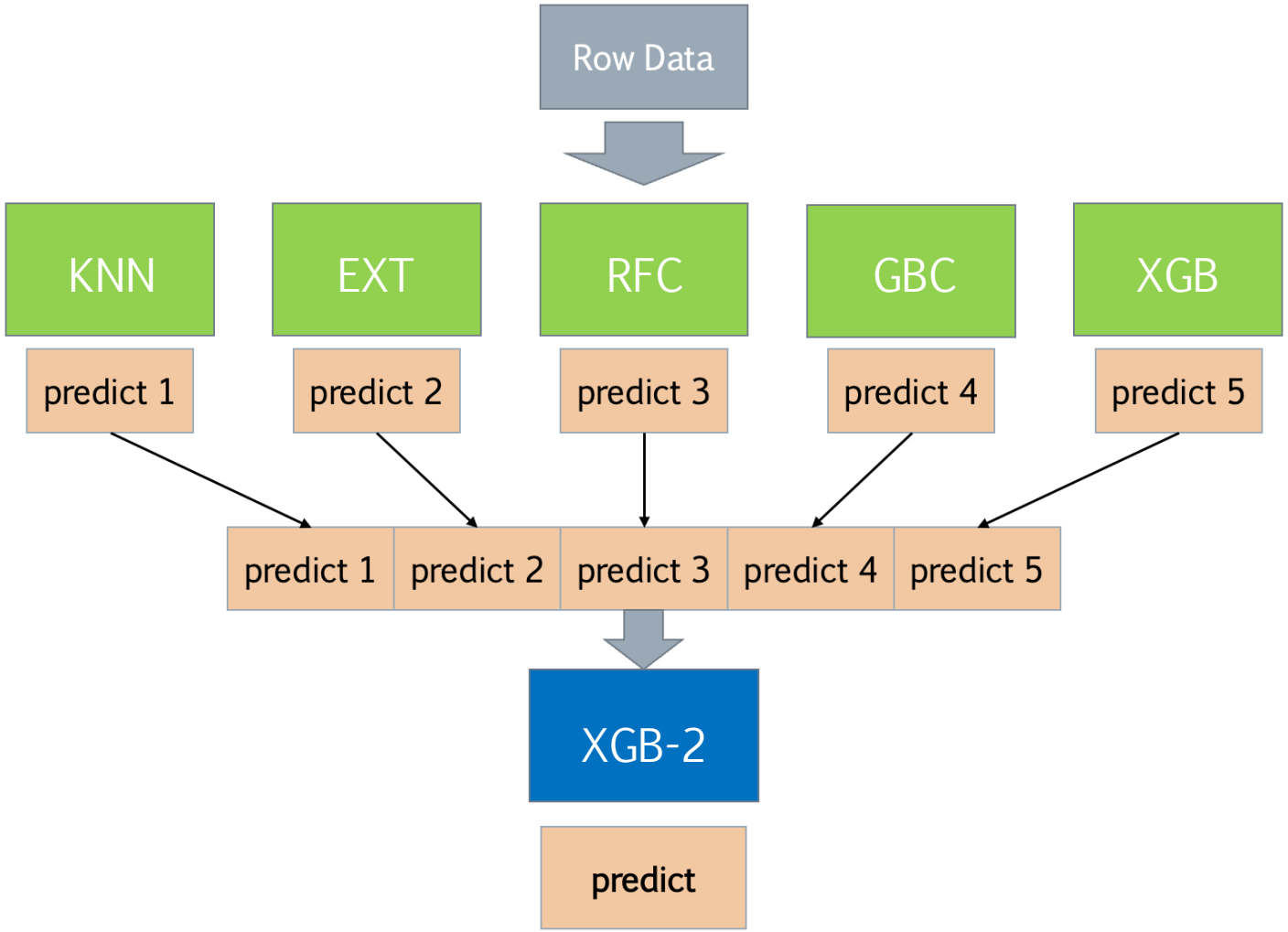
PR

Stacking model

Step 1:
Train multiple base models

Step 2:
Generate a new data using the predictions

Step 3:
Train a meta-model(XGB) on this new data



PR

Results and Conclusion

model	accuracy
KNN	0.742
RFC	0.854
EXT	0.854
GBC	0.837
XGB	0.831
Original	0.860

- Random forest(RFC) and Extra Trees(EXT) achieved the highest accuracy among base models
- The accuracy of Original model is the highest score of all

PR

Results and Conclusion

- We implemented a classification model that combines multiple models using the stacking method and verified its accuracy.
- Ensemble multiple models improved classification accuracy more than a single model.

- Next...
  - Tuning the optimal hyperparameters
  - implement stacking models with more than two layers