

ML-ENSEMBLE

Dongkyung IM

Department of Statistics

Chung-Ang University

January 17, 2022

1 ENSEMBLE

- Bagging
- Forests of randomized trees
- AdaBoost

ENSEMBLE

▷ **What is ENSEMBLE?**

- Combining the predictions of several base estimators
- To improve generalization (robustness) over a single estimator

▷ **Two families of ensemble methods**

- Averaging methods - Bagging , Forests of randomized trees
- Boosting methods - AdaBoost , Gradient Boosting

Bagging

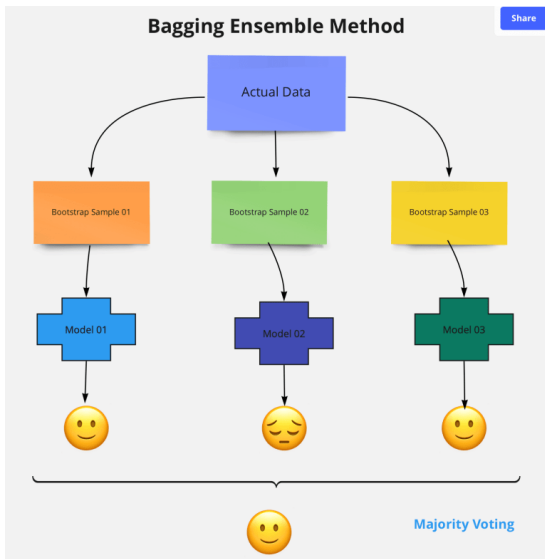
▷ **What is Bagging?**

- Bootstrap Aggregating
- Reducing the variance of a base estimator
- Working best with strong and complex models

▷ **Bagging methods come in many flavours.**

- Pasting
- Bagging
- Random Subspaces
- Random Patches

Bagging



Forests of randomized trees

▷ **Random Forests**

- Similar to Bagging
- Randomly select variables
- Diversity of the models

▷ **Extremely Randomized Trees**

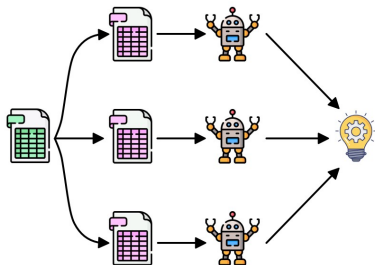
- Randomness goes one step further in the way splits are computed.
- Thresholds are drawn at random for each candidate feature.
- Reducing the variance of the model a bit more

▷ AdaBoost

- To fit a sequence of weak learners *on repeatedly modified data*
- applying weights to each of the training samples
- Reducing the bias of the model

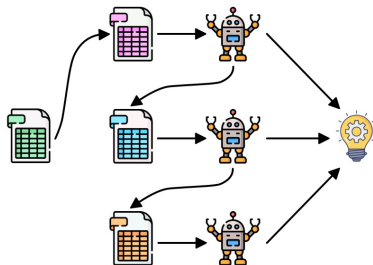
AdaBoost

Bagging



Parallel

Boosting



Sequential

THANK YOU

