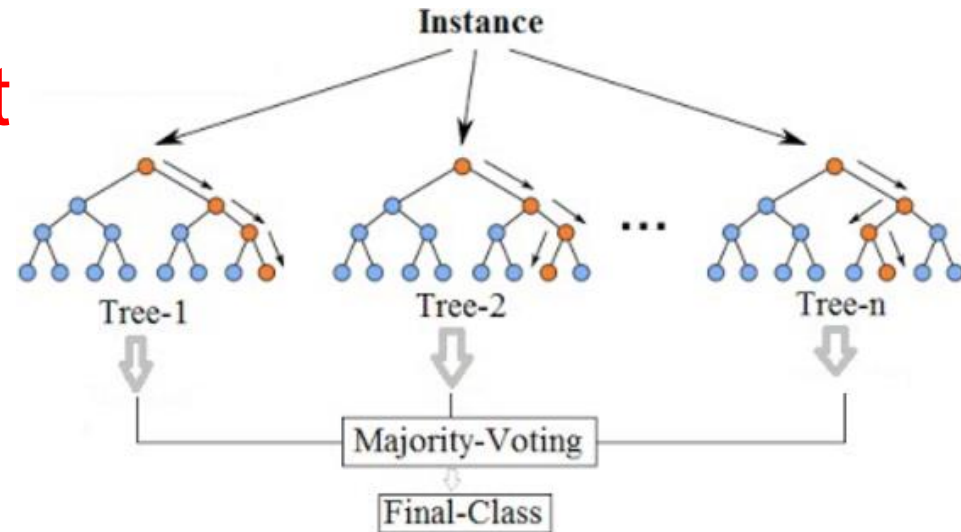# DAIS Lecture 15 – Ensemble Learning New Supplement

- Bagging using Decision Trees: Random Forest

- Gradient Boosting

# Random Forest



**Ensemble of decision trees**
with two bagging schemes:

- ordinary **bagging**

  - sample data with replacement to obtain different datasets

  - train one tree on each dataset

- **random subspace method** ("feature bagging")

  - use a random subset of the features for each candidate split

  ← this prevents different trees from
  choosing the same features for splits,
  which would make them highly correlated

# Gradient Boosting (I/II)

Idea: stage-wise learning, but use gradient

- generalizes to regression

■ Consider mean square error loss:

$$\text{L}_{\text{MSE}} = \frac{1}{\#data}\big(y - \text{H}_{\text{t}}(\text{x})\big)^2$$

teacher value → $y$

imperfect classifier at round t → $\text{H}_{\text{t}}(\text{x})$

$h_{t+1}$
residual: to be learnt by next classifier $h_{t+1}$

■ Next classifier $h_{t+1}$ accounts for the residual

■ Let this be minimized:

$$\frac{\partial}{\partial H_t}\text{L}_{\text{MSE}} = \frac{2}{\#data}\big(y - \text{H}_{\text{t}}(\text{x})\big) = \frac{2}{\#data}\text{h}_{t+1}(\text{x})$$

→ Each classifier sees a different loss

# Gradient Boosting (II/II)

- Gradient boosting trains an ensemble of (simple) models, each by gradient descent, and stage-wise

  - (whereas, ordinary stochastic gradient descent trains a single model)

- Next classifier model $h_{t+1}$ trained to correct the remaining errors of the previous models

  - (no use of weights, unlike in AdaBoost)

- Software: https://xgboost.ai

  - implements gradient boosted classification and regression trees