# 〈Tree ensembles - XGBoost〉

* Boosted trees intuition

⇒ Given training set of size m

For b = 1 to B :

- Using sampling with replacement to create a new training set of size m

  But instead of picking from all examples with equal (1/m) probability,

  make it more likely to pick examples that the previously trained trees missclassify.

- Train a decision tree on the new dataset.

i) sampling with replacement을 통해 새로운 training set 생성

ii) 새로운 training set을 통해 decision tree 생성

iii) original training set을 decision tree의 inputter로 해서 해당 decision tree가 original training example들을 잘 예측했는지 확인

iv) 다음 for loop에서는 똑같이 sampling with replacement을 통해 새로운 training set을 생성하되, iii)에서 제대로 예측하지 못한 example에 대해 higher chance를 부여하여 pick할수 있도록 함

⇒ "Boosting"

* XGBoost (eXtreme Gradient Boosting)

- Open Source
- Fast efficient implementation
- Good choice of default splitting criteria and criteria for when to stop splitting
- Built in regularization to prevent overfitting
- Highly competitive algorithm for machine learning competition