



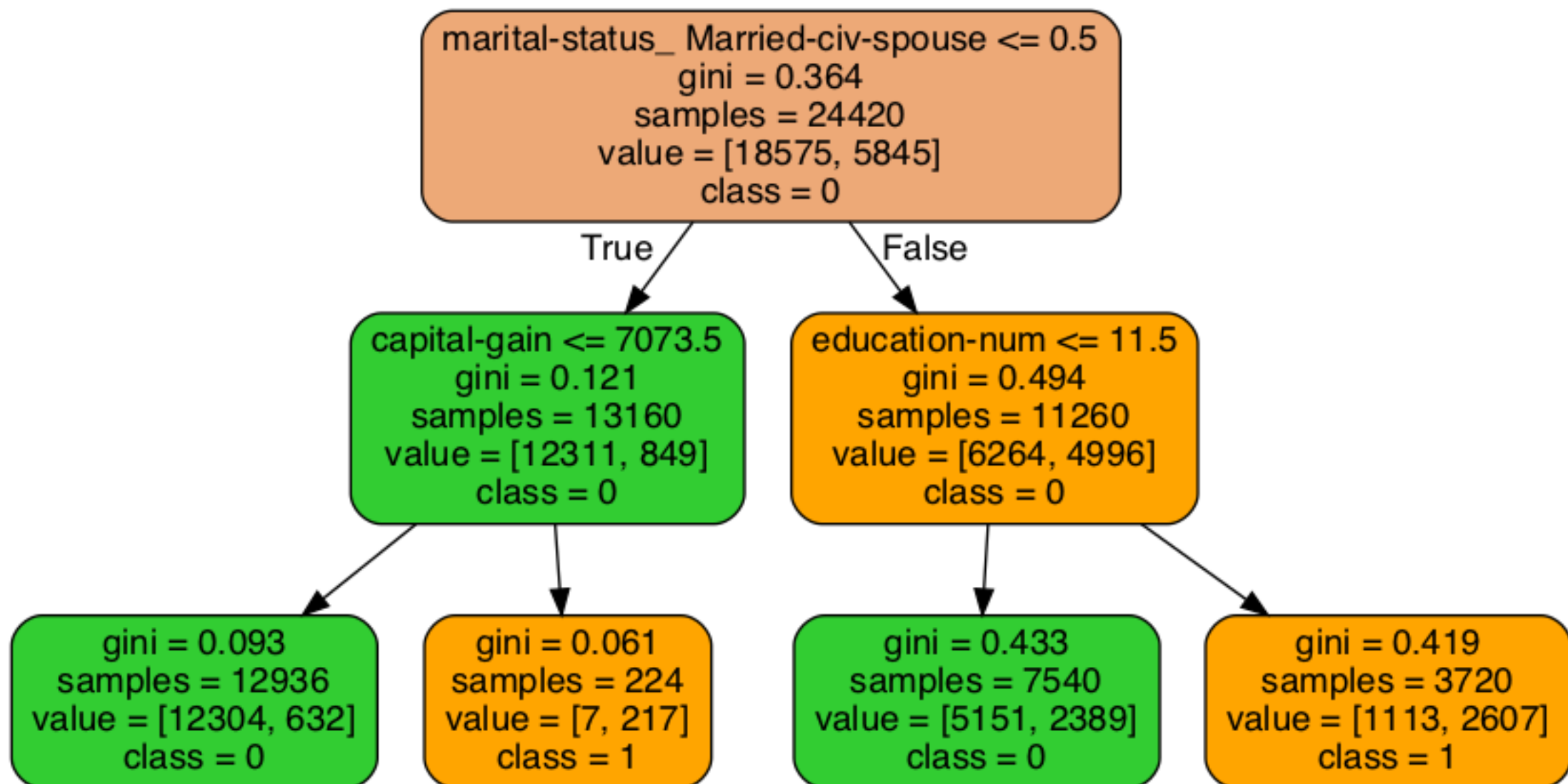
FROM DECISION TREES TO XGBOOST

DECISION TREES

- ▶ Like a game of 20 Questions
- ▶ Each round, the data is split in half
- ▶ Splits are made by determining which split will give the best score each round.
- ▶ The tree stops splitting when a desired level of accuracy has been reached.
- ▶ Since trees can keep going, they are prone to overfitting.
- ▶ Solutions? [Fine-tune trees](#) or [combine them into ensembles](#).

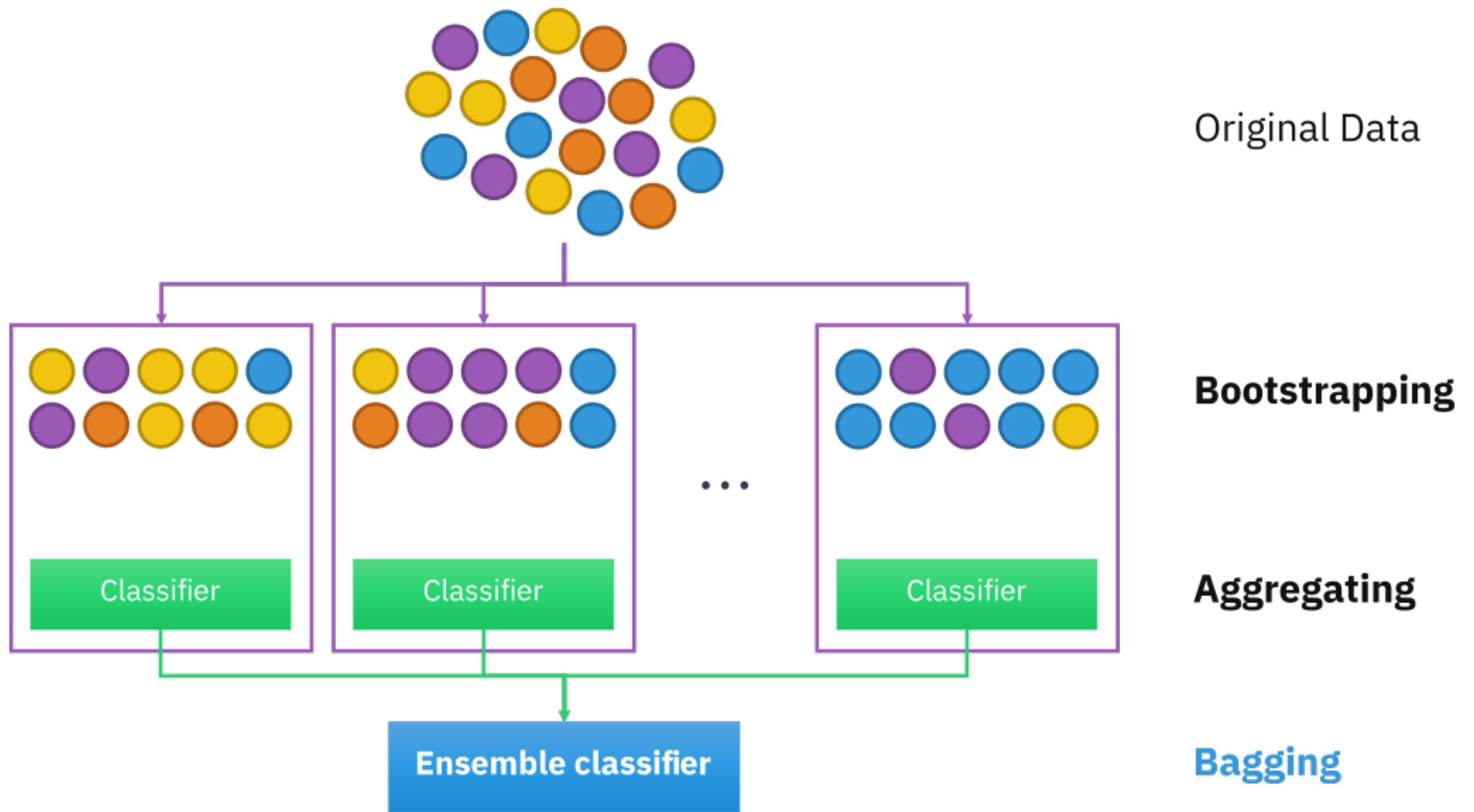
DECISION TREE - IMAGE

Census Dataset - max_depth=2



RANDOM FORESTS

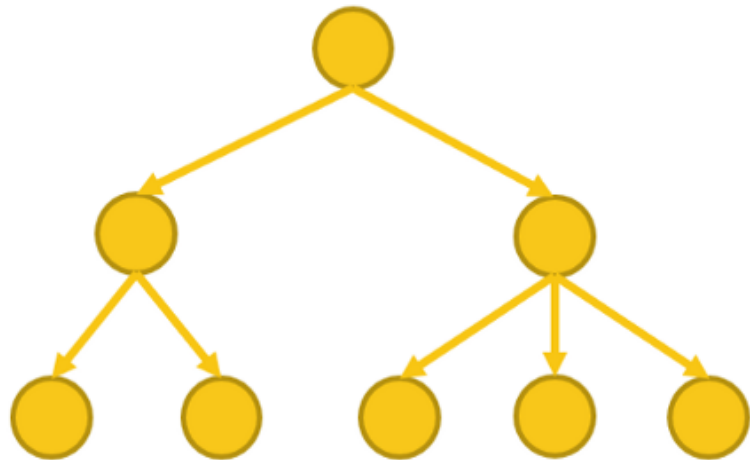
- ▶ Combine many Decision Trees (100 by Default).
- ▶ Trees are bootstrapped, meaning sampled with replacement.
- ▶ Randomness is involved with choosing the rows (bootstrapping) and the number of features (columns) to consider when making tree splits.
- ▶ Since the Decision Trees vary, they give different predictions.
- ▶ The Random Forest takes the average of all trees to make a final prediction.
- ▶ Random Forests use bagging: bootstrapping + aggregation.



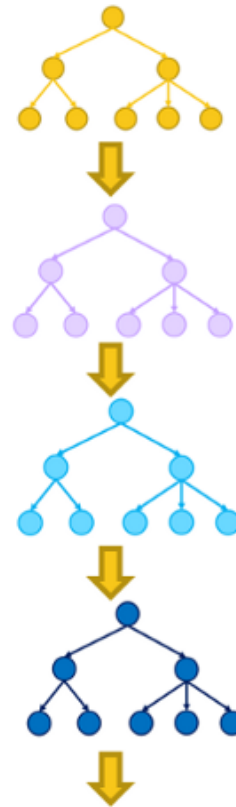
GRADIENT BOOSTING

- ▶ Random Forests are no better than individual trees.
- ▶ Random Forest trees do not improve with time.
- ▶ Gradient Boosting trains on the errors of each subsequent tree in the ensemble.
- ▶ Like Random Forests, Gradient Boosting combines many trees, but instead of bagging, they boost trees to make them stronger.
- ▶ Gradient Boosting can turn a weak learner into a strong learner.
- ▶ Boosting has been shown to be more effective than bagging overall.

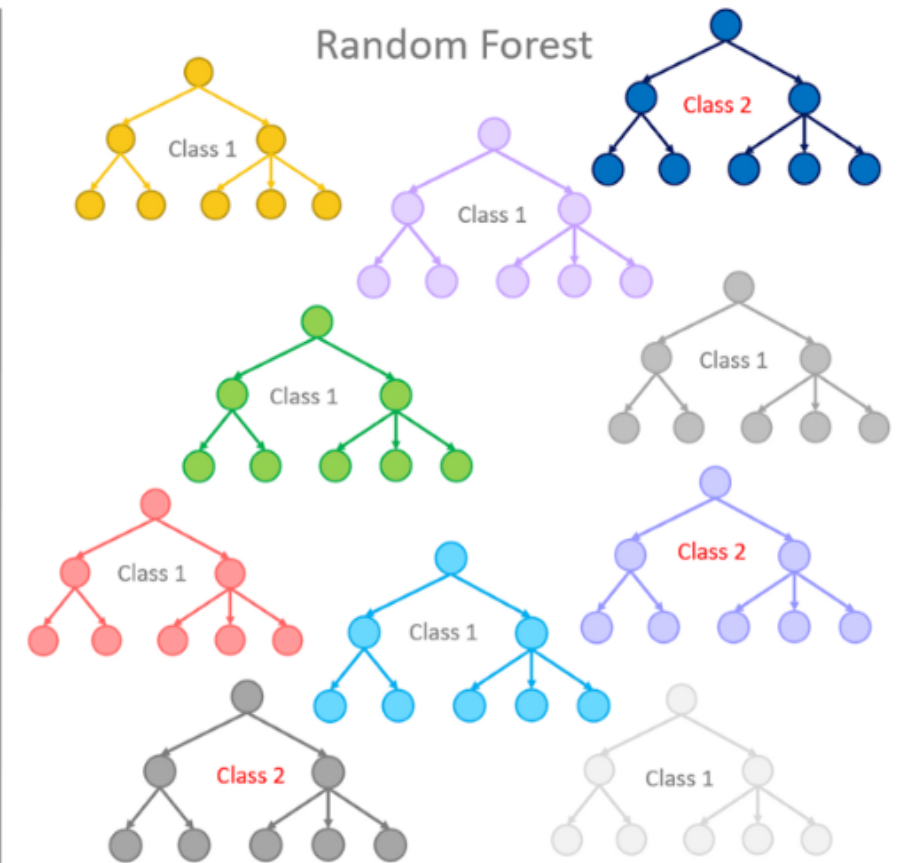
Single Decision Tree

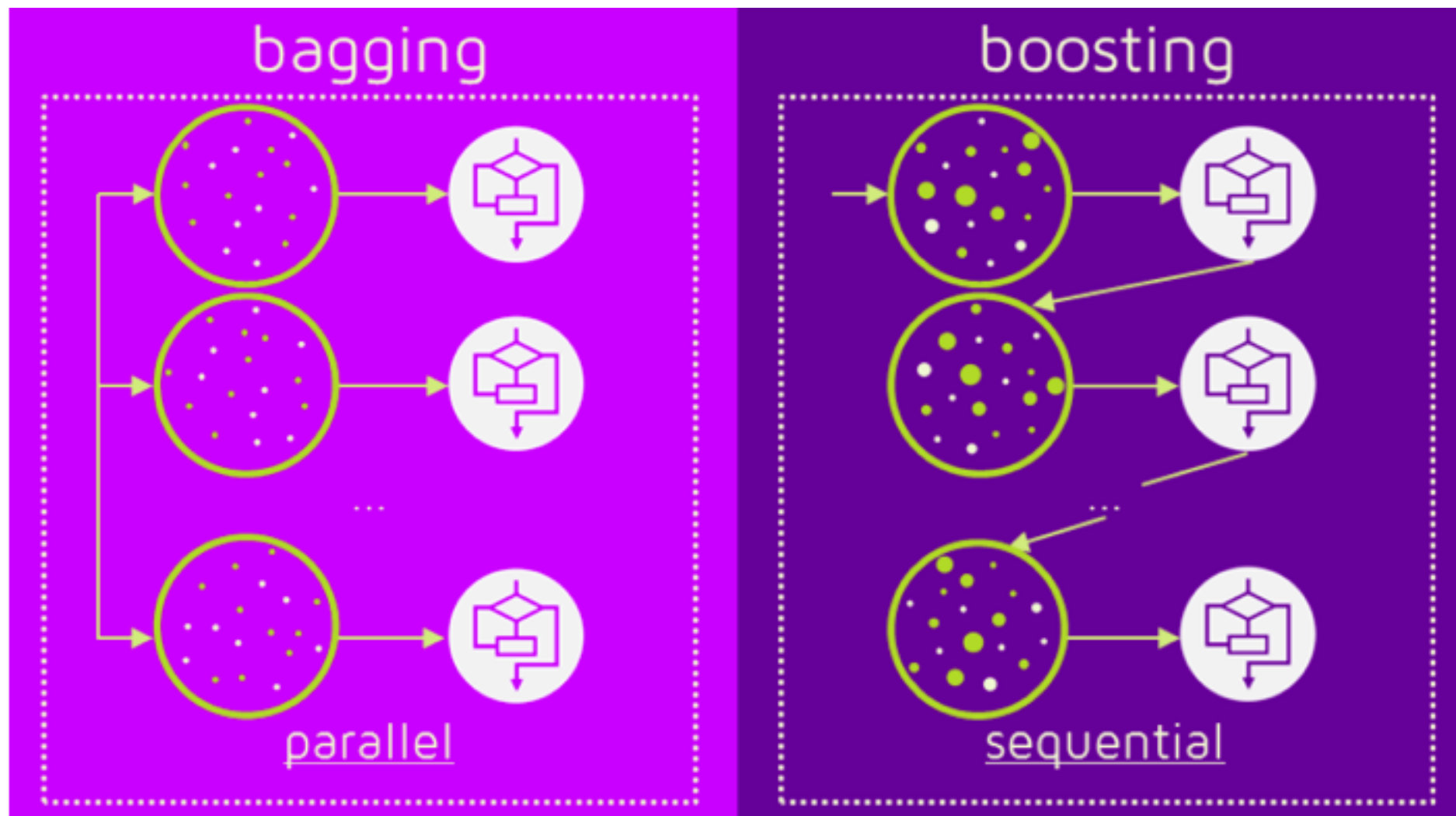


Gradient Boosted Trees



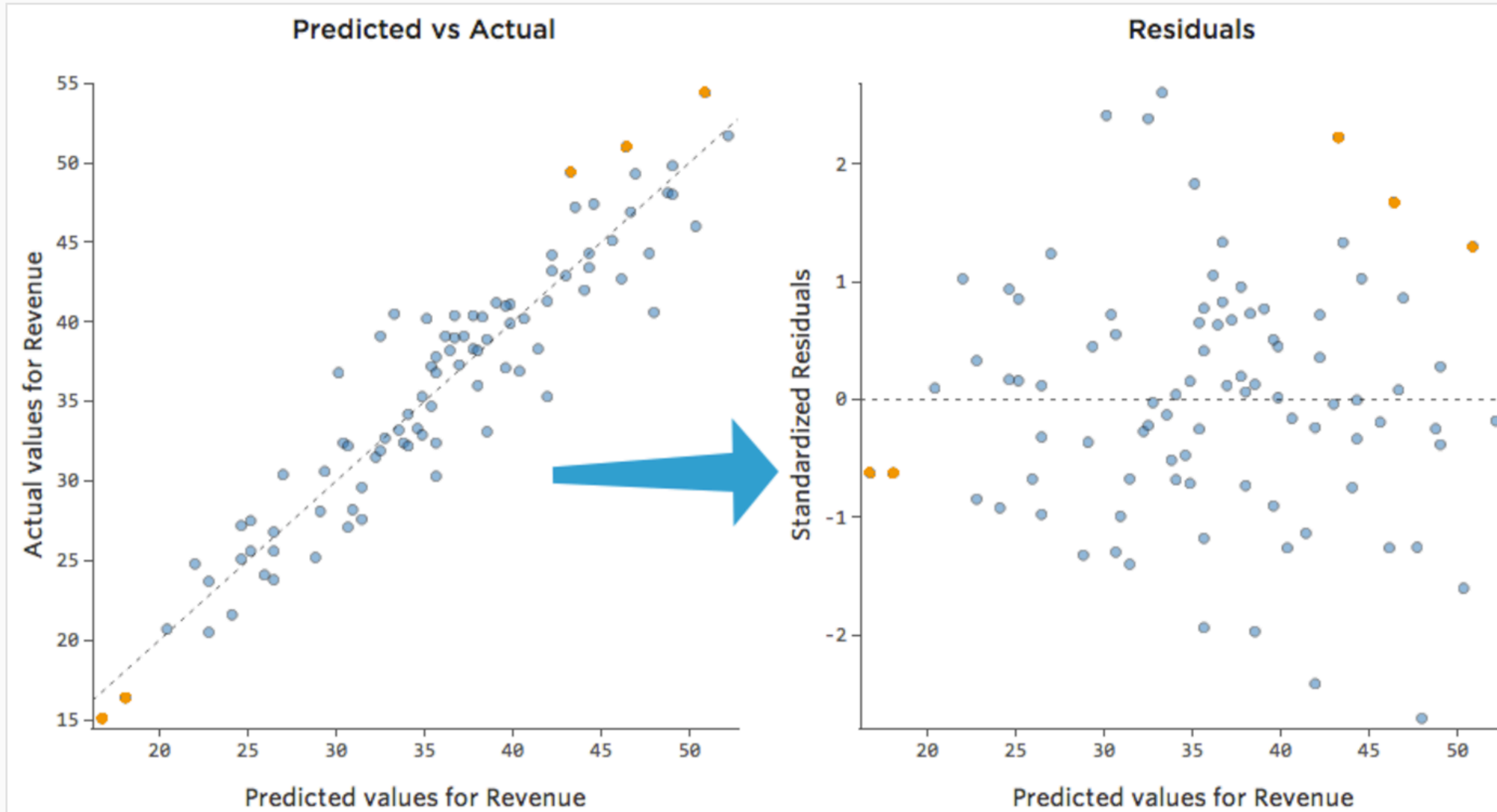
Random Forest





GRADIENT BOOSTING – RESIDUALS

- ▶ Gradient Boosting trains on the residuals.
- ▶ The residuals are the difference between the predictions and the test set.



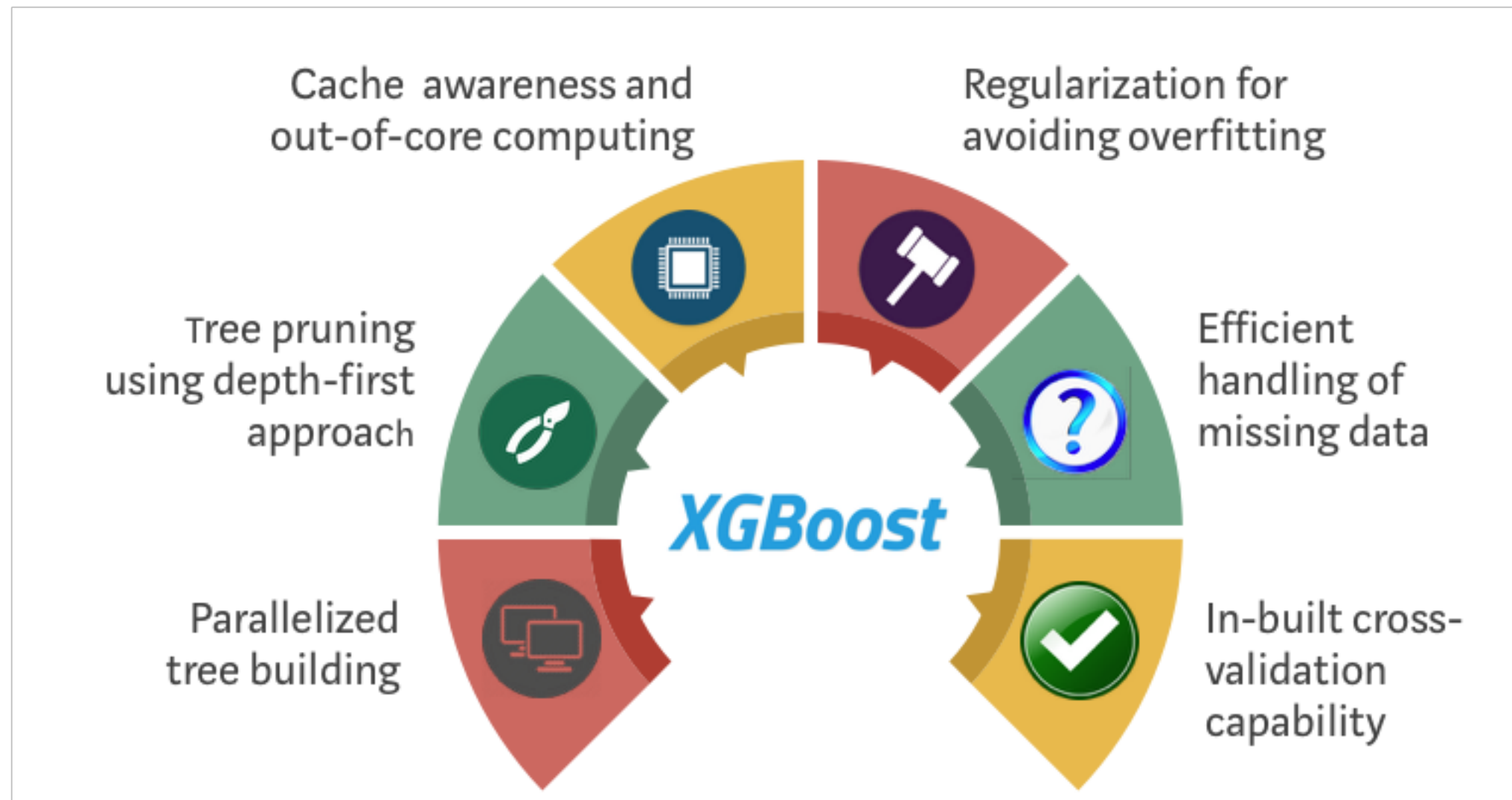
Note that we've colored in a few dots in orange so you can get the sense of how this transformation works.

GRADIENT BOOSTING – IMPROVEMENTS

- Can Gradient Boosting be faster?
- Can Gradient Boosting be more accurate?
- Can Gradient Boosting have more hyperparameters?
- Yes, with [XGBoost](#).

EXTREME GRADIENT BOOSTING – XGBOOST

- ▶ Faster than Gradient Boosting.
- ▶ Usually more accurate than Gradient Boosting.
- ▶ Includes unique hyperparameters to handle null values and imbalanced data.
- ▶ Includes alternative base learners beyond decision trees.
- ▶ Has won many Kaggle competitions.
- ▶ Arguably the best machine learning algorithm for handling tabular data (rows and columns of numerical and categorical data).



XGBOOST FACTS

- ▶ Famous for winning Higgs boson Kaggle competition to confirm existence of the Higgs boson.
- ▶ Scikit-learn wrapper new as of 2018 for XGBClassifier and XGBRegressor.
- ▶ XGBoost has their own Random Forests, XGBRFClassifier and XGBRFRegressor.
- ▶ XGBoost uses sparse matrices to speed up computations. Performs 10 times faster than Gradient Boosting.
- ▶ XGBoost takes advantage of hardware/software components like cache and parallel processing to maintain an edge in speed.

XGBOOST LINKS

- ▶ XGBoost Documentation - <https://xgboost.readthedocs.io/en/latest/>
- ▶ XGBoost Tutorial - <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>
- ▶ XGBoost original paper - <https://arxiv.org/pdf/1603.02754.pdf>
- ▶ Youtube link from XGBoost - https://www.youtube.com/watch?v=Vly8xGnNiWs&t=2136s&ab_channel=DataScience.LA
- ▶ XGBoost hyperparameters - <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>
- ▶ XGBoost book github page - <https://github.com/PacktPublishing/Hands-On-Gradient-Boosting-with-XGBoost-and-Scikit-learn->



XGBOOST FOUNDER – TIANQI CHEN



