Supervised, Unsupervised and Reinforcement Learning in Finance

Week 1: Supervised Learning

Tree methods: Random Forests

Igor Halperin

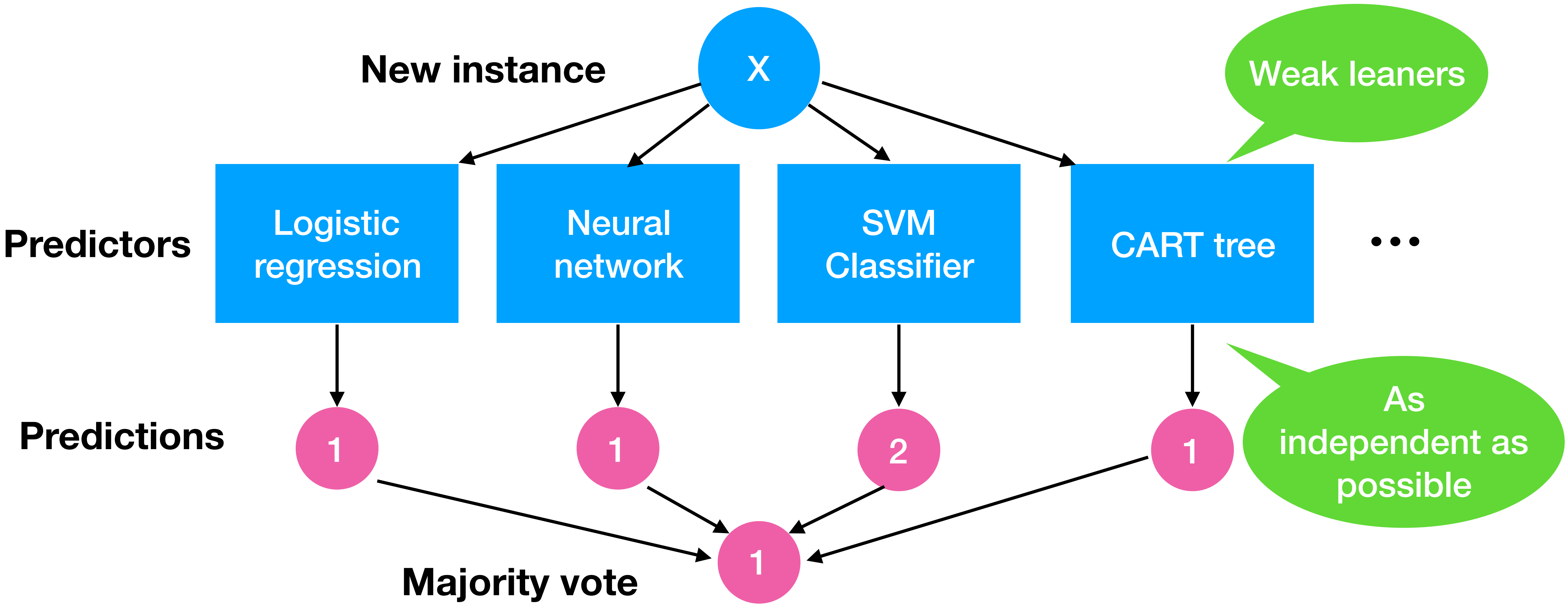
NYU Tandon School of Engineering, 2017

Ensemble methods for trees

We can combine (aggregate) predictions made by many classifiers, each trained on the same input data.

This is called **Ensemble Learning.**

How it can work: train many classifiers, then use a majority vote to define the final class for an instance.



Ensemble learning: a coin example

Assume we have a biased coin with the head probability of 51%.

Toss the coin $N = 1000$ times.

1. Each trial produces a **weak learner** with the probability of $p_H = 0.51$ to get the right answer.

2. Construct a **strong learner** by taking a majority vote among weak learner

$$P\left(N_H > N_T\right) = P\left(N_H > \frac{N}{2}\right) = 1 - P\left(N_H \leq \frac{N}{2}\right) \simeq 73\%$$

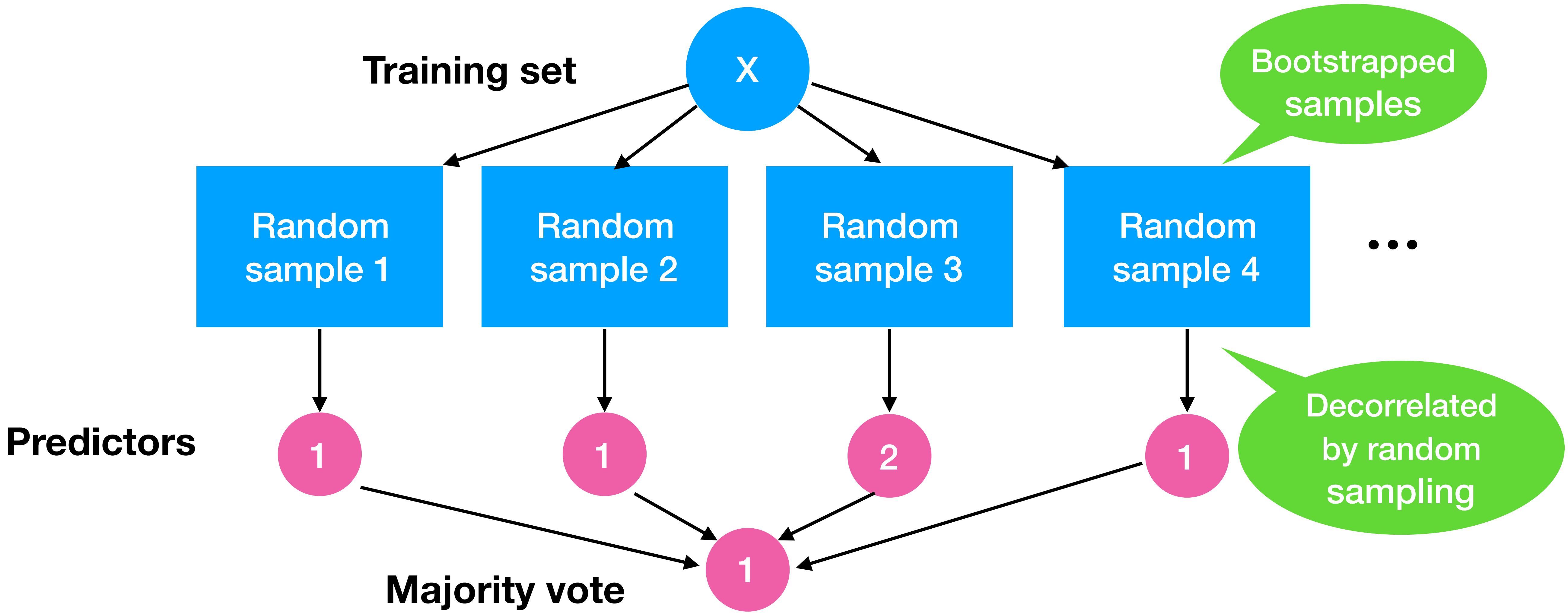
3. The strong learner uses the **law of large numbers** to make a prediction that is better than the best predictor in its set of weak learners.

Bagging

Bagging = bootstrap aggregation (Breiman, 1996)

Bagging = Ensemble Learning with weak learners obtained from the same algorithm

How it works: train many trees on random sub-sets of the input data, then use a majority vote to define the final class for an instance.

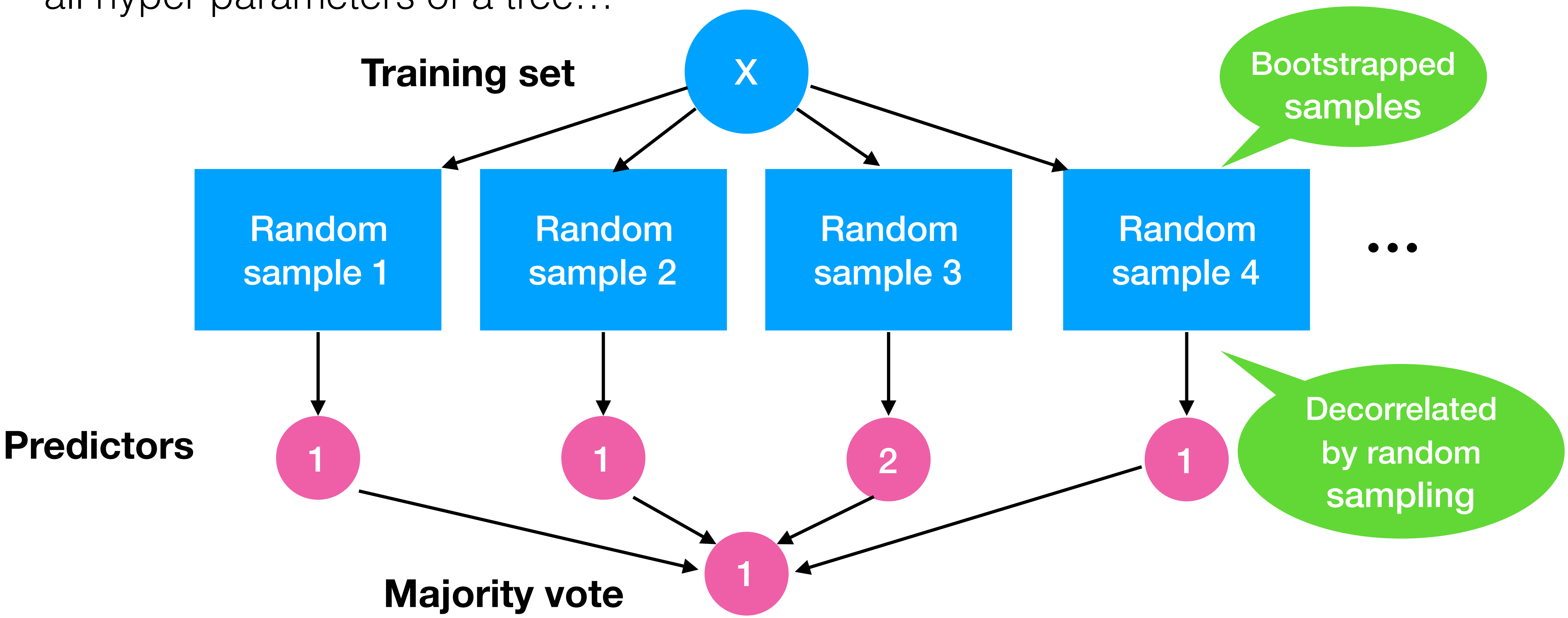


Random Forests

Random Forest (Breiman, 2001): de-correlate base (weak) learners by randomly sampling features and subsets of a train dataset, use bagging for predictions.

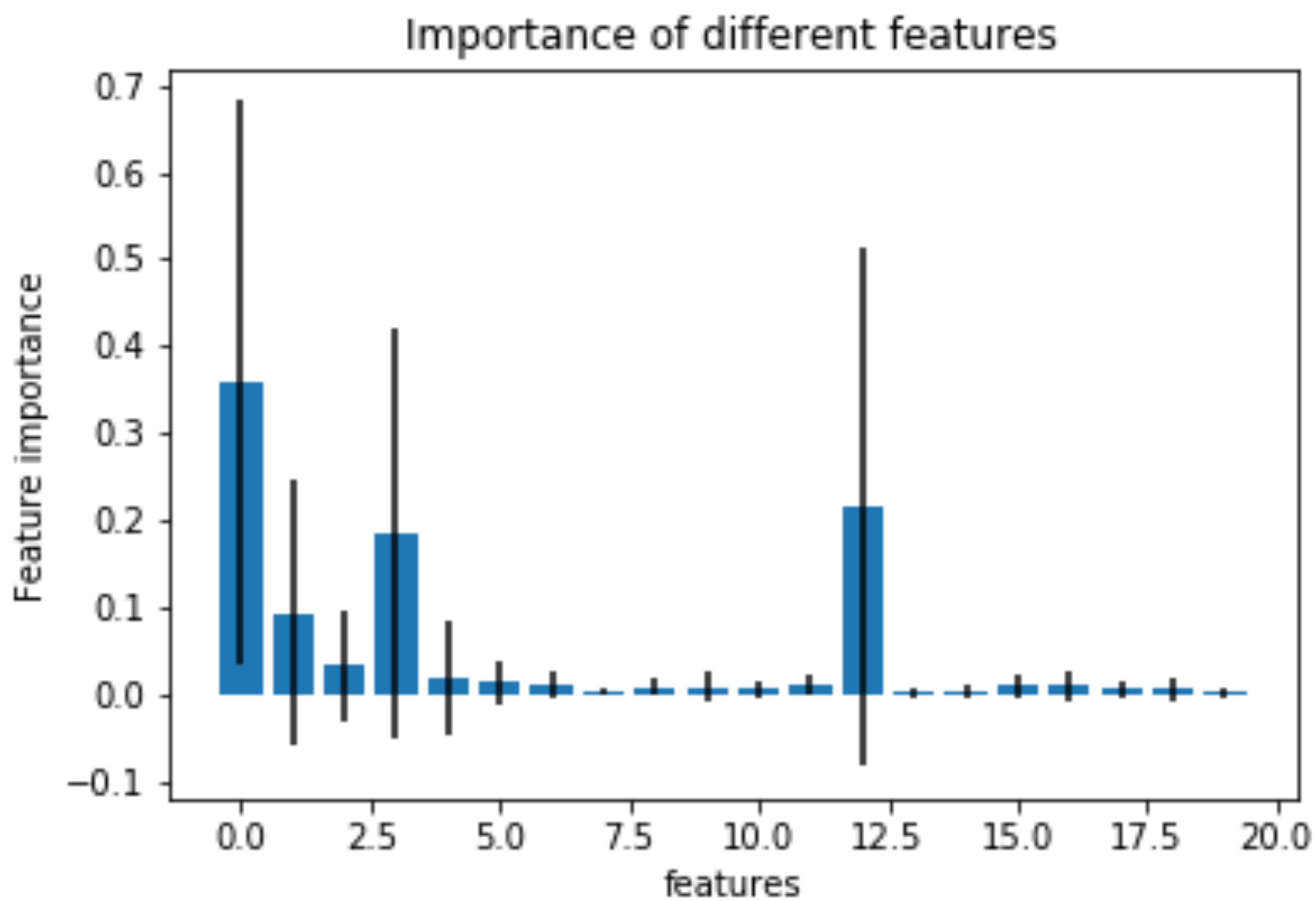
Produces lower variance estimators than a tree

Hyper-Parameters of a Random Forest:
- num_estimators
- all hyper-parameters of a tree…



Feature importance in Random Forest

Important of feature is given by an average depth at which at appears across all trees in a Random Forest, plus a standard deviation of this depth:



	Importance	Std
log_TA	0.350043	0.329071
NI_to_TA	0.094549	0.156222
Equity_to_TA	0.030421	0.035744
NPL_to_TL	0.133095	0.192915
REO_to_TA	0.025687	0.077842
ALLL_to_TL	0.011211	0.034182
core_deposits_to_TA	0.010999	0.013458
brokered_deposits_to_TA	0.001025	0.003567
liquid_assets_to_TA	0.009313	0.011609
loss_provision_to_TL	0.011886	0.021393
NIM	0.011005	0.013408
assets_growth	0.003713	0.006769
FDIC_assessment_base_n	0.271504	0.320079

Control question

Select all correct answers

1. Weak learners in ensemble methods should be as independent of each other as possible.

2. A strong learner in ensemble learning is the best learner among weak learners.

3. Performance of the strong performer is better than performance of the best weak learning, as ensemble learning relies on the law of large numbers to reduce variance of the final estimator.

4. Random forest algorithm creates de-correlated weak learning by random sampling of subsets of features and subsets of data points.

5. As it is based on randomization, a Random Forest estimator typically has higher variance than a single tree estimator.

Correct answers: 1, 3, 4.