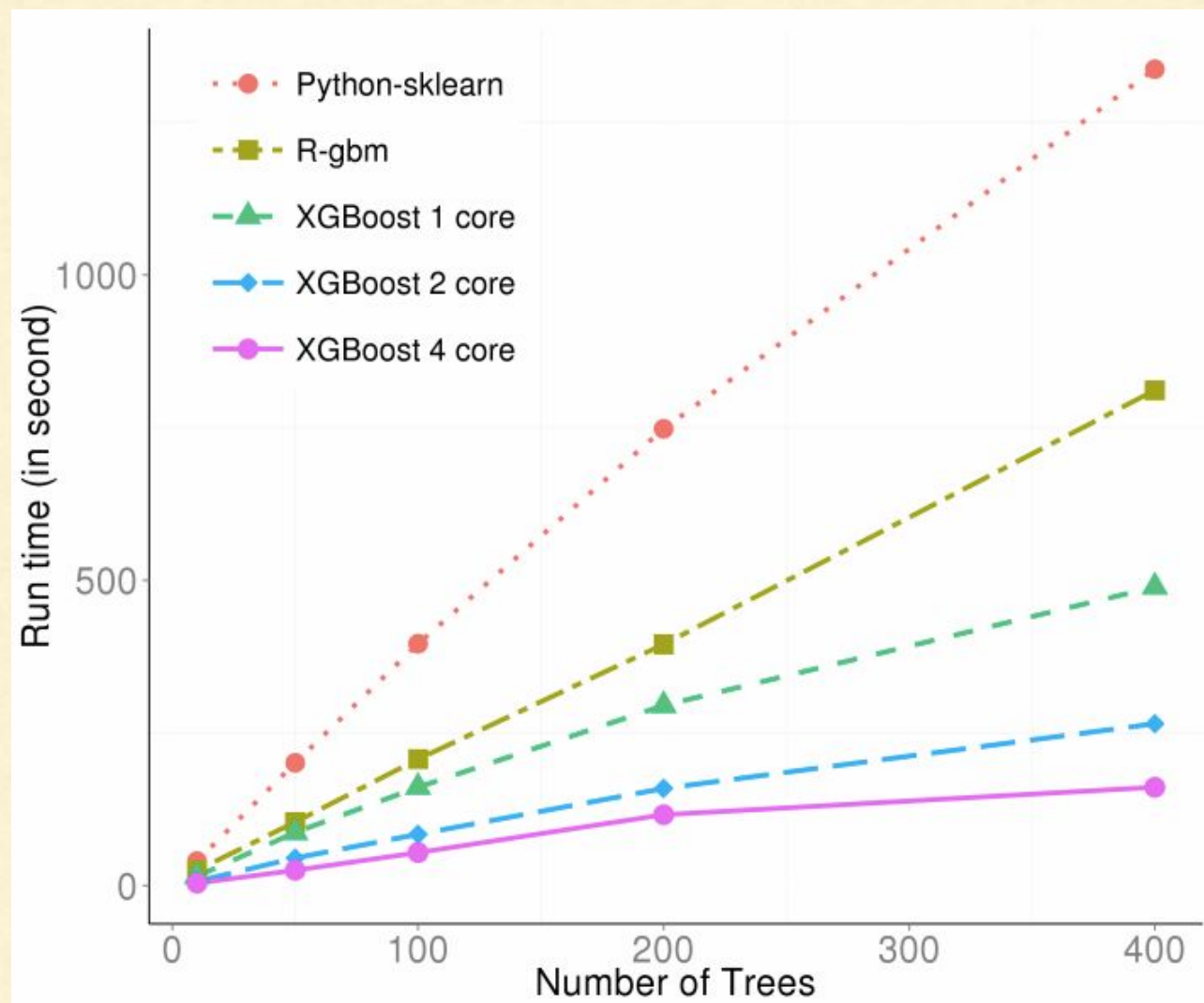




XGBOOST: A MACHINE LEARNING TOOL

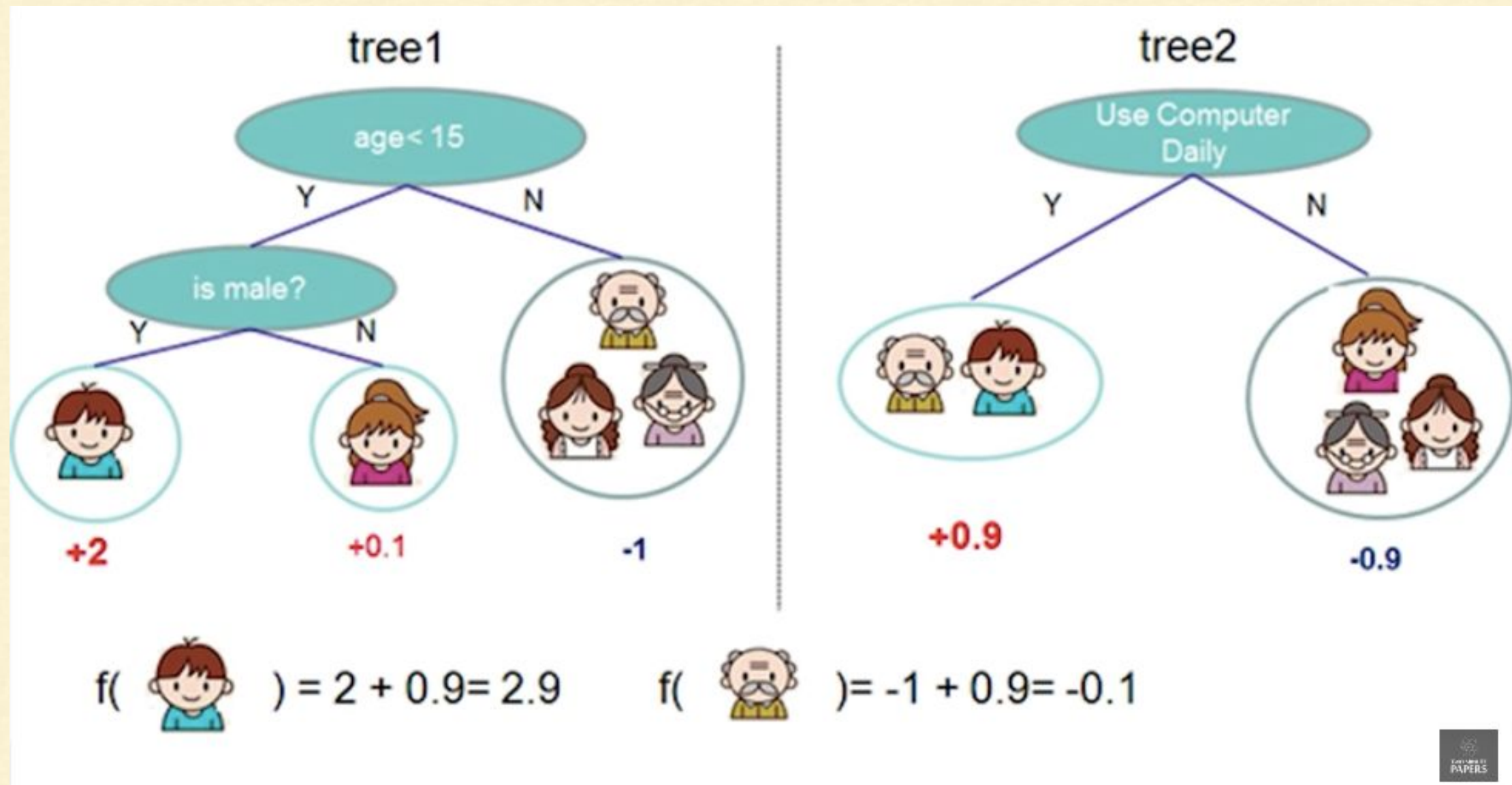
By Celine Seghossian

WHAT IS XGBOOST?



- . A machine learning tool for extreme gradient boosting
- . Used with structured data sets to predict outcomes based off of multiple characteristics of the data
- . Faster than other gradient boosting algorithms
- . Has helped win many Kaggle competitions :)

DECISION TREES & BOOSTING



Boosting combines many decision trees to get accurate models.
XGBoost uses gradient descent to minimize loss when adding new models.

OUR PROBLEM

- . Data: “Adult Data Set”, census data on income from 1996
 - . Given census data, can we predict whether a person makes over \$50K a year?
 - . An advantage of XGBoost over Neural Networks is that we can see which attributes the computer uses to make its decisions!
 - . Which factors (e.g. age, sex, education) most heavily correlate to a person’s income in 1996?
-

WEBLINKS

- . Decision Trees and Boosting: https://www.youtube.com/watch?v=0Xc9LIb_HTw
 - . Graph - XGBoost vs. other methods: <http://proceedings.mlr.press/v42/chen14.pdf>
 - . Main XGBoost Tutorial: <https://jessesw.com/XG-Boost/>
 - . Adult Data Set: <https://archive.ics.uci.edu/ml/datasets/Adult>
 - . Redwood Trees Image: <https://www.shutterstock.com/g/flipov>
-