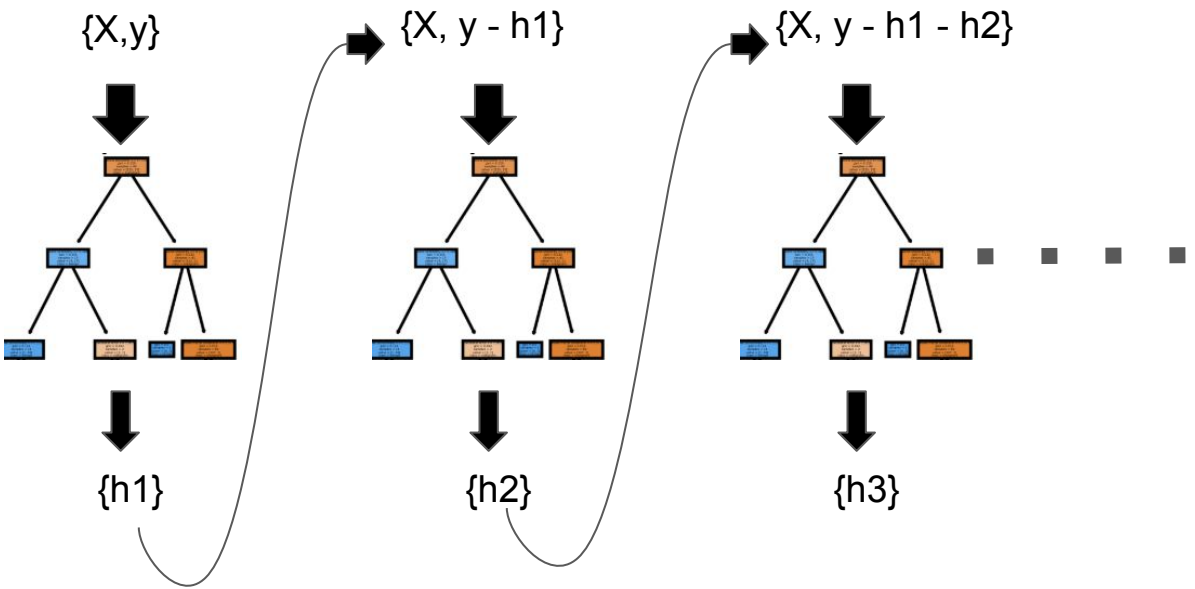
GBM vs XGboost

What are Gradient Boosting Machines?

- Most of the answer is in the name: “Gradient” and “Boosting”.
- **Boosting** is referring to the ensemble of boosted weak learners.
- **Gradient** part is referring to minimising or a **Descent** for a loss function.
- The ensemble of Weak learners used are typically **decision trees**.
- GBM supports both regression and classification.

Gradient Boosting Machines

Dataset : {X, y}



Final Model = sum / boost of the predictors

h1 can start off as base rate of mean of y or another simple regression tree

Loss function and Gradient Descent intuition

Loss function or E(Error) using MSE formula in this case

$$E = (y - h)^2/2$$

We take partial derivative of loss function to find min

$$\partial E/\partial h = (y - h).(-1)$$

Gradient descent algorithm minimizes function by iteratively moving in the direction of steepest descent as defined by the negative gradient.

$$h = h - \alpha.(\partial E/\partial h)$$

$$h = h - \alpha.\ (y - h).(-1)$$

$$\mathbf{h = h + \alpha\ (y - h)}$$

Residuals in Regression and Classification

- Regression Residuals are typically difference between y - ŷ.
- Classification Residuals are based on “Log of Odds” and Probability.
- More complex to calculate that Regression Residuals.
- The log of the odds of y are calculated and converted into probability.
- Residual calculations then can be done.
- The residuals are then transformed back to log odds again.

$$\text{Probability} = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$$

XGBoost : eXtreme Gradient Boosting

- An enhanced implementation of GBM.
- Known as a more regularized model to control over-fitting.
- The name xgboost, refers to the engineering goal to push the limit of computations resources for boosted tree algorithms.
- Use of second partial derivatives for more informative gradient descent.
- Uses more “Advanced” decision Trees than GBM .
- Uses Regularization (L1 & L2), which improves model generalization.
- Training is very fast and can be parallelized over CPUs

XGBoost : eXtreme Gradient Boosting

- Can be Distributed across computing clusters such as Hadoop.
- High Flexibility, allows user defined custom optimization & evaluation criteria.
- XGBoost has an in-built routine to handle missing values.
- More advanced tree pruning algorithm than GBM
- Has built in cross validation option
- + more

References and additional information:

https://www.gormanalysis.com/blog/gradient-boosting-explained/

http://mccormickml.com/2014/03/04/gradient-descent-derivation

https://www.youtube.com/watch?v=3CC4N4z3GJc

https://youtu.be/OtD8wVaFm6E

https://explained.ai/gradient-boosting/faq.html

https://www.quora.com/What-is-the-difference-between-the-R-gbm-gradient-boosting-machine-and-xgboost-extreme-gradient-boosting