Extreme Gradient Boosting

Brandon Croom

Agenda

- Boosting Refresher
- Gradient Boosting
- XGBoost
- Computational Complexity
- GBMs vs XGBoost
- Other Implementations
- Summary

Boosting Refresher

- **Boosting Models**
  - Form of ensemble modeling using a decision tree in which the trees are formed sequentially.
  - Each sequential tree will use the error of the prior tree as its target and seek to minimize the error of the target variable
  - Are generally quite accurate in their predictions but run the risk of overfitting

Boosting Refresher

- Boosting Notation:
  - Initial Model: $F_0(x)$
  - New Model: $h_1(x)$

  - Boosted Version:

$$F_1(x) \leftarrow F_0(x) + h_1(x)$$

  - Generalized Notation:

$$F_m(x) \leftarrow F_{m-1}(x) + h_m(x)$$

Agenda

- Boosting Refresher
- Gradient Boosting
- XGBoost
- Computational Complexity
- GBMs vs XGBoost
- Other Implementations
- Summary

Gradient Boosting - Enhancements

- Enhances boosting process through:
  - Learning rate
  - Sample size

Gradient Boosting – The math

- Notation
  - Define: $F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$
  - Compute the gradient of the loss function

$$r_{im} = -\alpha\left[\frac{\partial(L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}, \text{ where } \alpha \text{ is the learning rate}$$

  - Generalized Notation:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Agenda

- Boosting Refresher
- Gradient Boosting
- XGBoost
- Computational Complexity
- GBMs vs XGBoost
- Other Implementations
- Summary

XGBoost

- Started as a research project by Tianqi Chen for the Distributed Machine Learning Community in 2014
- Created initially as a terminal application
- Became well known in ML competitions after winning the Higgs Machine Learning Challenge
- Enhanced version released by Tianqi Chen and Carlos Guestrin in 2016

XGBoost - Enhancements

- Algorithmic
  - Tree Pruning
  - Sparsity Aware Split Finding
  - Continued Training
- Systematic:
  - Parallelization
  - Cache Aware
  - Distributed Computing
  - Out-of-Core Computing
- Flexibility
  - Customized Objective Function
  - Customized Evaluation Metric
- Cross-validation
  - Built-in Cross-validation

XGBoost - The Math

- XGBoost Objective Function:

$$\frac{\partial L(y, f^{(m-1)}(x) + f_m(x))}{\partial f_m(x)} = 0$$

XGBoost - The Math

- Taylor expansion:

$$L(y, f^{(m-1)}(x) + f_m(x)) \approx L(y, f^{(m-1)}(x)) + g_m(x) f_m(x) + \frac{1}{2} h_m(x) f_m(x)^2,$$

- Hessian:

$$h_m(x) = \frac{\partial^2 L(Y, f(x))}{\partial f(x)^2}_{f(x)=f^{(m-1)}(x)}.$$

XGBoost - The Math

- Rewritten loss function:

$$L(f_m) \approx \sum_{i=1}^{n} [g_m(x_i) f_m(x_i) + \frac{1}{2} h_m(x_i) f_m(x_i)^2] + const.$$

$$\propto \sum_{j=1}^{T_m} \sum_{i \in R_{jm}} [g_m(x_i) w_{jm} + \frac{1}{2} h_m(x_i) w_{jm}^2].$$

- Consolidated loss function:

$$L(f_m) \propto \sum_{j=1}^{T_m} [G_{jm} w_{jm} + \frac{1}{2} H_{jm} w_{jm}^2].$$

XGBoost - The Math

- Optimal Weight:

$$w_{jm} = -\frac{G_{jm}}{H_{jm}}, j = 1, ..., T_m.$$

- New Loss function:

$$L(f_m) \propto -\frac{1}{2}\sum_{j=1}^{T_m}\frac{G_{jm}^2}{H_{jm}}.$$

Gain Function:

$$Gain = \frac{1}{2}[\frac{G_{jmL}^2}{H_{jmL}} + \frac{G_{jmR}^2}{H_{jmR}} - \frac{G_{jm}^2}{H_{jm}}]$$

$$= \frac{1}{2}[\frac{G_{jmL}^2}{H_{jmL}} + \frac{G_{jmR}^2}{H_{jmR}} - \frac{(G_{jmL} + G_{jmR})^2}{H_{jmL} + H_{jmR}}].$$

XGBoost – The Math

- Final Loss Function:

$$L(f_m) \propto \sum_{j=1}^{T_m}[G_{jm}w_{jm} + \frac{1}{2}H_{jm}w_{jm}^2] + \gamma T_m + \frac{1}{2}\lambda\sum_{j=1}^{T_m} w_{jm}^2 + \alpha\sum_{j=1}^{T_m}|w_{jm}|$$

$$= \sum_{j=1}^{T_m}[G_{jm}w_{jm} + \frac{1}{2}(H_{jm}+\lambda)w_{jm}^2 + \alpha|w_{jm}|] + \gamma T_m,$$

$$w_{jm} = \begin{cases} -\frac{G_{jm}+\alpha}{H_{jm}+\lambda} & G_{jm} < -\alpha, \\ -\frac{G_{jm}-\alpha}{H_{jm}+\lambda} & G_{jm} > \alpha, \\ 0 & else. \end{cases}$$

$$Gain = \frac{1}{2}[\frac{T_\alpha(G_{jmL})^2}{H_{jmL}+\lambda} + \frac{T_\alpha(G_{jmR})^2}{H_{jmR}+\lambda} - \frac{T_\alpha(G_{jm})^2}{H_{jm}+\lambda}] - \gamma$$

$$T_\alpha(G) = \begin{cases} G+\alpha & G < -\alpha, \\ G-\alpha & G > \alpha, \\ 0 & else. \end{cases}$$

Agenda

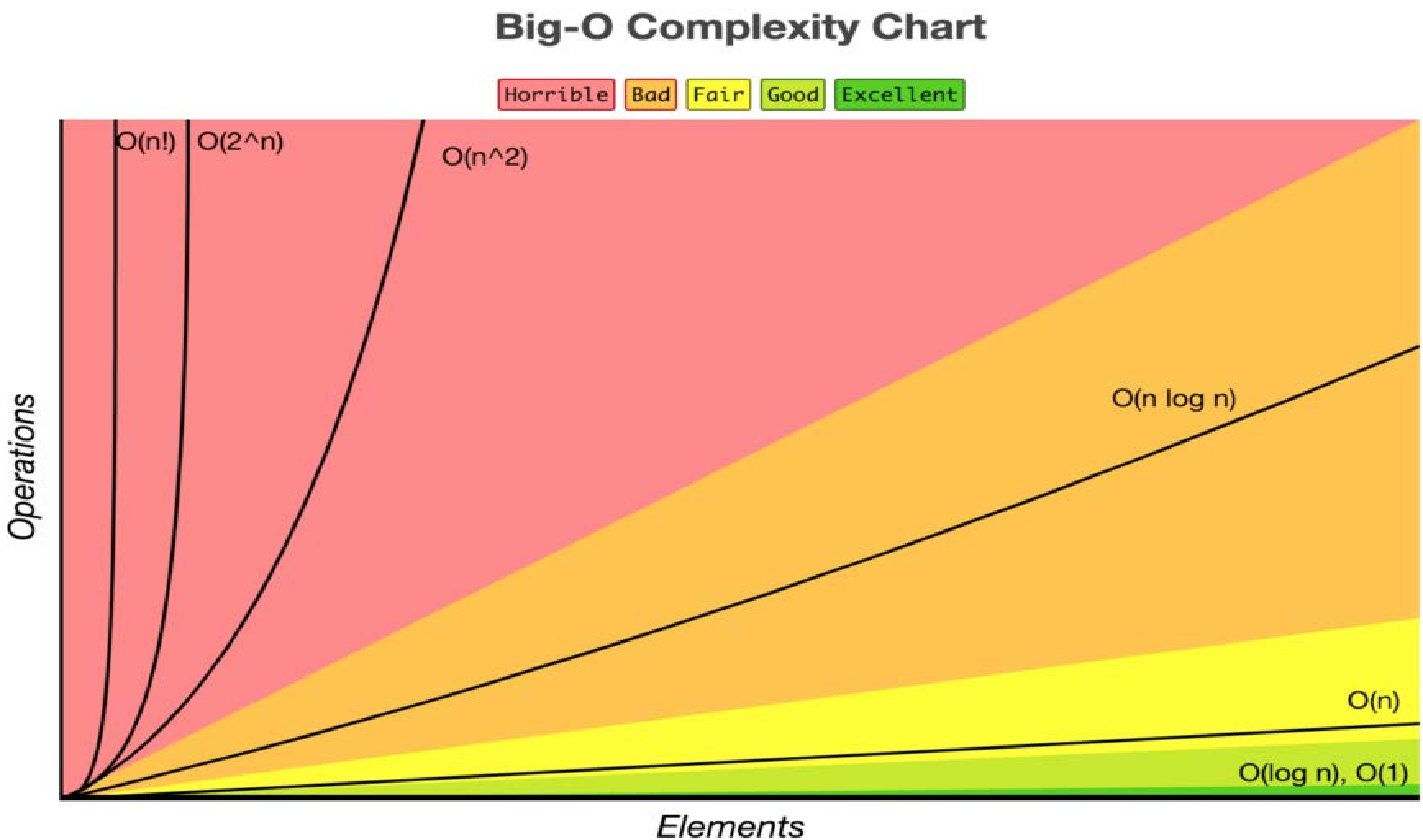
- Boosting Refresher
- Gradient Boosting
- XGBoost
- Computational Complexity
- GBMs vs XGBoost
- Other Implementations
- Summary

Computational Complexity

- Big O Notation – used to classify algorithms
  - Based on run time or space requirement growth as input size grows

- ML Big O Notation can be difficult

Computational Complexity



Computational Complexity

- Gradient Boosting

  $O(Knp)$

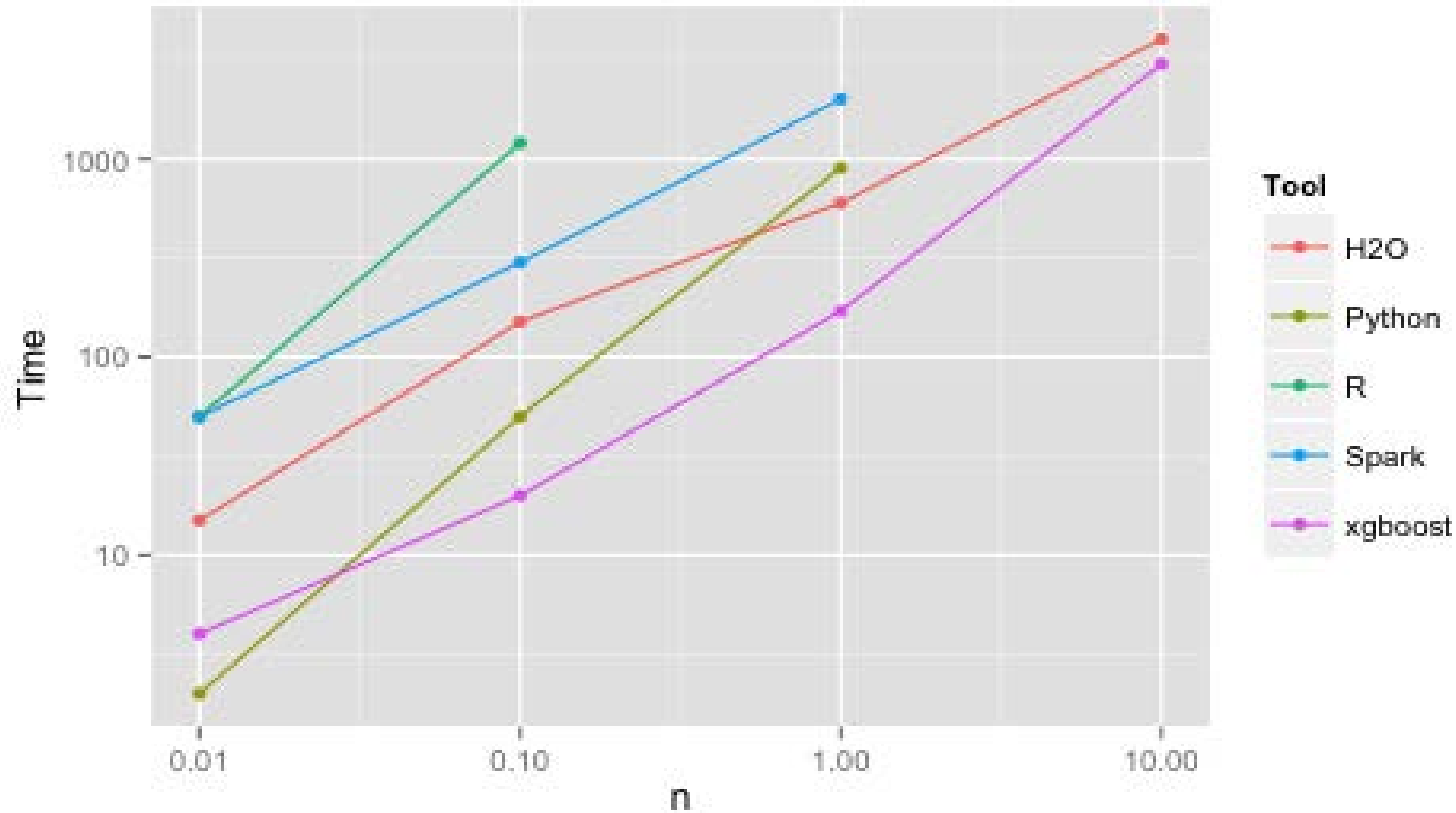
- Original Sparse Greedy Algorithm

  $O(Kd\|\mathbf{x}\|_0 \log n)$

- XGBoost Big O

  $O(Kd\|\mathbf{x}\|_0 + \|\mathbf{x}\|_0 \log B)$

Computational Complexity



Agenda

- Boosting Refresher
- Gradient Boosting
- XGBoost
- Computational Complexity
- GBMs vs XGBoost
- Other Implementations
- Summary

GBM vs. XGBoost

- GBM has a broader application
  - Both need to calculate gradient at current estimate
  - XGBoost also needs to calculate a hessian
    - Requires objective function to be twice differentiable
  - GBM only requires a differentiable loss function

GBM vs. XGBoost

- XGBoost is faster
  - GBM weight calculation is the average value of the gradients
  - XGBoost weight calculation is the sum of gradients scaled by the sum of hessians

$$w_{jm} = \begin{cases} -\frac{G_{jm}}{n_{jm}} & GBM, \\ -\frac{G_{jm}}{H_{jm}} & XGBoost. \end{cases}$$

GBM vs. XGBoost

- XGBoost provides more regularization options:
  - Can regularize L1 and L2
  - Penalize on number of leaf nodes

$$Gain = \begin{cases} \frac{G_{jmL}^2}{n_{jmL}} + \frac{G_{jmR}^2}{n_{jmR}} - \frac{G_{jm}^2}{n_{jm}} & GBM, \\ \frac{1}{2}\left[\frac{T_\alpha(G_{jmL})^2}{H_{jmL}+\lambda} + \frac{T_\alpha(G_{jmR})^2}{H_{jmR}+\lambda} - \frac{T_\alpha(G_{jm})^2}{H_{jm}+\lambda}\right] - \gamma & XGBoost. \end{cases}$$

GBM vs. XGBoost

- XGBoost provides more randomization:
  - GBM provide only one level of column sampling
  - XGBoost provides two levels of column sample:
    - ByTree
    - ByLevel

Agenda

- Boosting Refresher
- Gradient Boosting
- XGBoost
- Computational Complexity
- GBMs vs XGBoost
- Other Implementations
- Summary

Other Implementations

- LightGBM
  - Framework released by Microsoft in 2017
  - Highly efficient, scalable framework
  - Supports multiple algorithms
  - Fast and memory efficient
  - Uses Gradient based One-Side Sampling
  - Handles Categorical features efficiently

- CatBoost
  - Framework released by Yandex Technology in 2017
  - Open source framework for gradient boosting decision trees
  - Handles Categorical features efficiently
  - Does not support sparse matrices
  - Longer runtimes on data sets with large number of numerical features

Agenda

- Boosting Refresher
- Gradient Boosting
- XGBoost
- Computational Complexity
- GBMs vs XGBoost
- Other Implementations
- Summary

Summary

- Provided a background on boosting techniques
- Deep dive into XGBoost
- Evaluated computational complexity
- Compared differences in GBMs and XGBoost
- Brief overview of other alternatives

References

- https://opendatascience.com/xgboost-is-machine-learnings-captain-america/
- https://arxiv.org/pdf/1603.02754.pdf
- https://medium.com/sfu-cspmp/xgboost-a-deep-dive-into-boosting-f06c9c41349
- https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/
- https://en.wikipedia.org/wiki/XGBoost
- https://homes.cs.washington.edu/~tqchen/2016/03/10/story-and-lessons-behind-the-evolution-of-xgboost.html
- read://https_machinelearningmastery.com/?url=https%3A%2F%2Fmachinelearningmastery.com%2Fgentle-introduction-xgboost-applied-machine-learning%2F
- https://towardsdatascience.com/boosting-algorithm-xgboost-4d9ec0207d
- https://medium.com/hackernoon/boosting-algorithms-adaboost-gradient-boosting-and-xgboost-f74991cad38c
- https://medium.com/hackernoon/gradient-boosting-and-xgboost-90862daa6c77
- https://opendatascience.com/gradient-boosting-and-xgboost/
- https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf
- https://github.com/szilard/benchm-ml
- https://towardsdatascience.com/catboost-vs-light-gbm-vs-xgboost-5f93620723db
- https://medium.com/@hanishsidhu/whats-so-special-about-catboost-335d64d754ae
- https://www.microsoft.com/en-us/research/project/lightgbm/
- https://catboost.ai/

Video Link

- https://vimeo.com/442735566