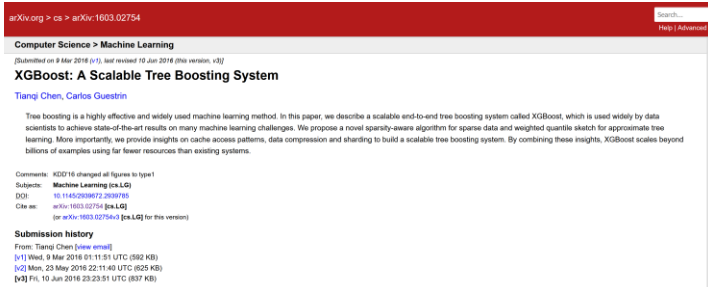
eXtreme Gradient Boosting

Xiao Lang

Department of Mechanics and Maritime Sciences
Division of Marine Technology
Gothenburg, SWEDEN

What is XGBoost ?

- eXtreme Gradient Boosting = XGBoost
- XGBoost is machine learning library like numpy, tensorflow, pytorch
- XGBoost has dominated machine learning hackathons and competitions



arXiv.org > cs > arXiv:1603.02754

Search...

Help | Advanced

Computer Science > Machine Learning

[Submitted on 9 Mar 2016 (v1), last revised 10 Jun 2016 (this version, v3)]

XGBoost: A Scalable Tree Boosting System

Tianqi Chen, Carlos Guestrin

Tree boosting is a highly effective and widely used machine learning method. In this paper, we describe a scalable end-to-end tree boosting system called XGBoost, which is used widely by data scientists to achieve state-of-the-art results on many machine learning challenges. We propose a novel sparsity-aware algorithm for sparse data and weighted quantile sketch for approximate tree learning. More importantly, we provide insights on cache access patterns, data compression and sharding to build a scalable tree boosting system. By combining these insights, XGBoost scales beyond billions of examples using far fewer resources than existing systems.

Comments:	KDD'16 changed all figures to type1
Subjects:	**Machine Learning (cs.LG)**
DOI:	10.1145/2939672.2939785
Cite as:	arXiv:1603.02754 **[cs.LG]**
	(or arXiv:1603.02754v3 **[cs.LG]** for this version)

Submission history

From: Tianqi Chen [view email]

[v1] Wed, 9 Mar 2016 01:11:51 UTC (592 KB)

[v2] Mon, 23 May 2016 22:11:40 UTC (625 KB)

[v3] Fri, 10 Jun 2016 23:23:51 UTC (837 KB)

Department of Mechanics and Maritime Sciences
Division of Marine Technology

Evolution of XGBoost

- **DT → Boosting → GBDT → XGBoost**

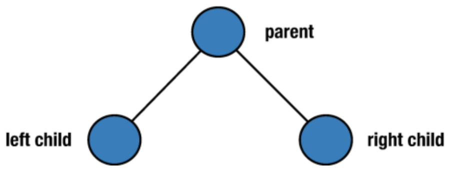
Time	Model	Original article
1986	DT	Induction of **Decision Trees**
1995	Boosting	A Decision-Theoretic Generalization of On-line learning and am Application to **Boosting**
2001	GBDT	Greedy Function Approximation: A **Gradient Boosting** Machine
2016	XGBoost	**XGBoost**: A Scalable Tree Boosting System

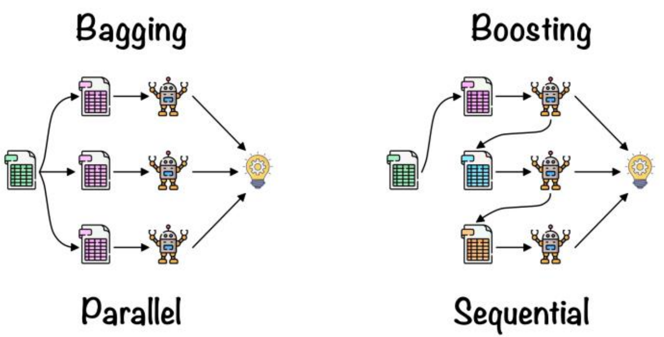
Evolution of XGBoost

- Decesion tree: ID3, C4.5, CART (binary tree)
- Boosting: one of ensemble learning method
- Gradient boosting decision tree
- XGBoost: CART (binary tree)

➢ classification

➢ regression

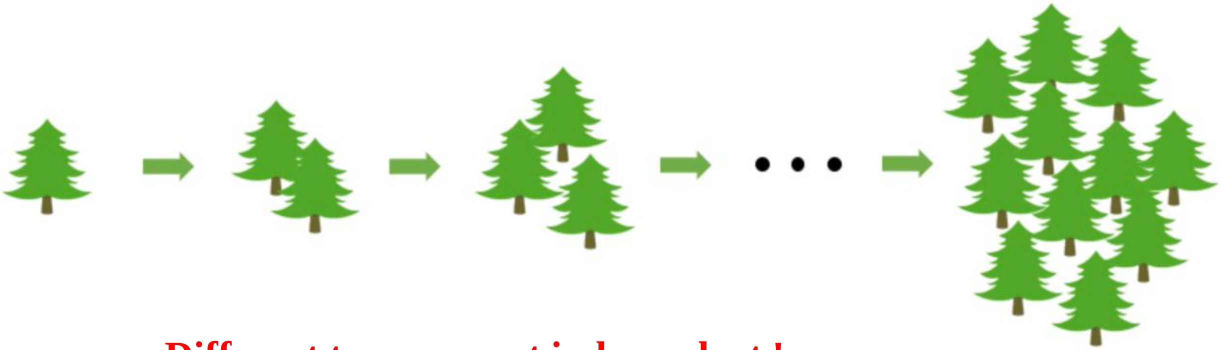




Department of Mechanics and Maritime Sciences
Division of Marine Technology

GBDT & XGBoost similarity

- **Boosting: establish tree (weak evaluator) one by one and accumulation**



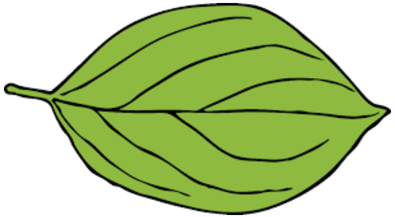
Different trees are not independent !

GBDT & XGBoost difference

- Decision tree

Regression

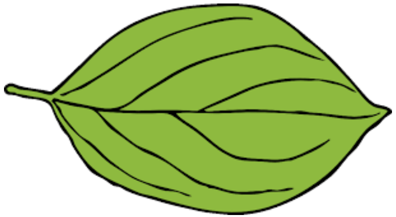
Leaf node: average



Sample	Actual
1	0.3
2	0.2
3	1.5
4	0.8
5	0.6
Prediction	0.68

Classification

Leaf node: majority



Sample	Actual
1	0
2	1
3	0
4	0
5	1
Prediction	0

Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

GBDT & XGBoost difference

- **Gradient boosting decision tree**

$$\hat{y}_i^{(k)} = \sum_{k}^{K} \gamma_k \boxed{h_k(x_i)}$$

Average or majority

- **XGBoost**

$$\hat{y}_i^{(k)} = \sum_{k}^{K} \boxed{f_k(x_i)}$$

Prediction score / leaf weight

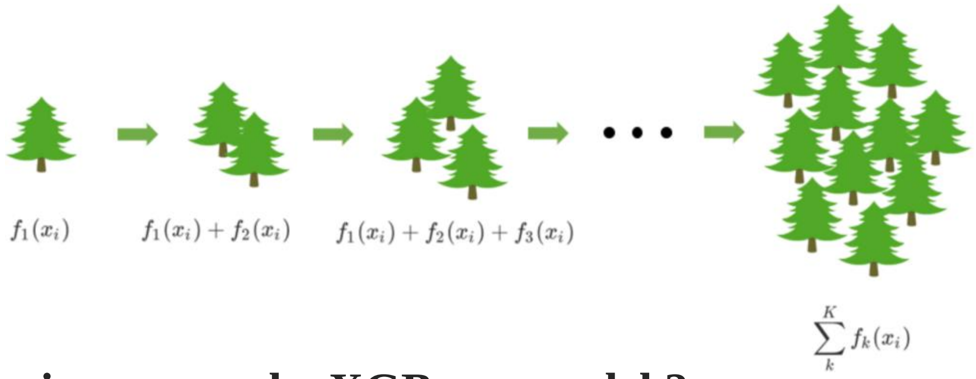
XGBoost parameter

class xgboost.XGBRegressor *(**max_depth**=3, **learning_rate**=0.1, **n_estimators**=100, **silent**=True, **objective**='reg:linear', **booster**='gbtree', **n_jobs**=1, **nthread**=None, **gamma**=0, **min_child_weight**=1, **max_delta_step**=0, **subsample**=1, **colsample_bytree**=1, **colsample_bylevel**=1, **reg_alpha**=0, **reg_lambda**=1, **scale_pos_weight**=1, **base_score**=0.5, **random_state**=0, **seed**=None, **missing**=None, **importance_type**='gain', **kwargs)*

- **Ensemble: n_estimators, learning_rate (eta)…**
- **Weak evaluator: max_depth, gamma, reg_alpha, reg_lambda…**
- **Application process: n_jobs…**

XGBoost parameter - - ensemble

- **n_estimator: number of trees to be grown**

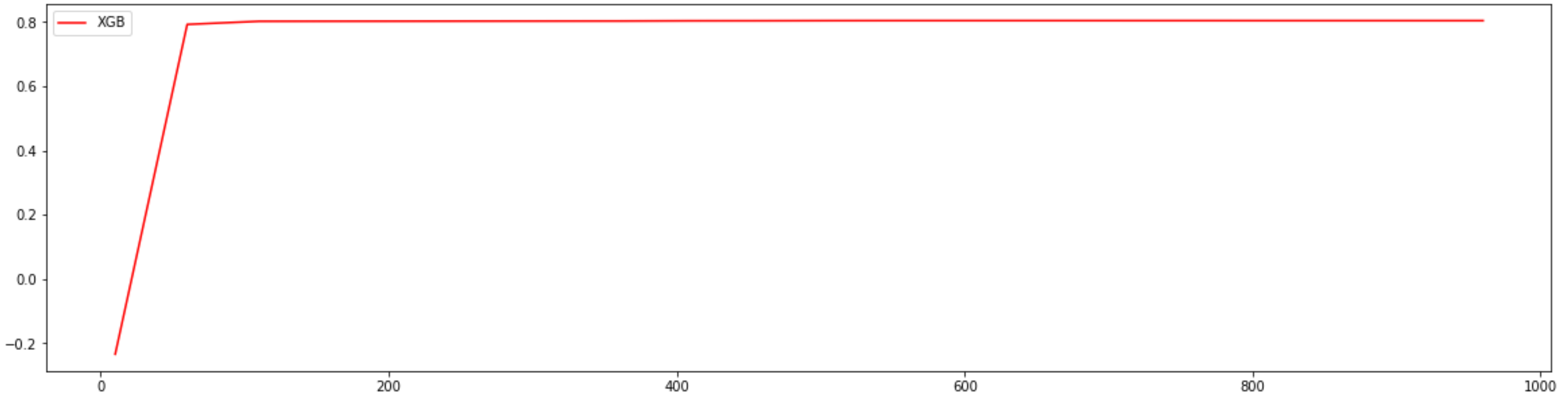


- **Influence of n_estimator on the XGBoost model ?**

Department of Mechanics and Maritime Sciences
Division of Marine Technology

XGBoost parameter - - ensemble

- **Influence of n_estimator on the XGBoost model ?**



XGBoost parameter - - ensemble

- **learning_rate (eta): step size shrinkage used in tree grow**

Logistic regression	⇨	**best fit**	⇨	**obejctive function**	⇨	$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \alpha * d_{ki}$
Gradient boosting	⇨	**best prediction**	⇨	**obejctive function**	⇨	$\hat{y_i}^{(k+1)} = \hat{y_i}^{(k)} + f_{k+1}(x_i)$

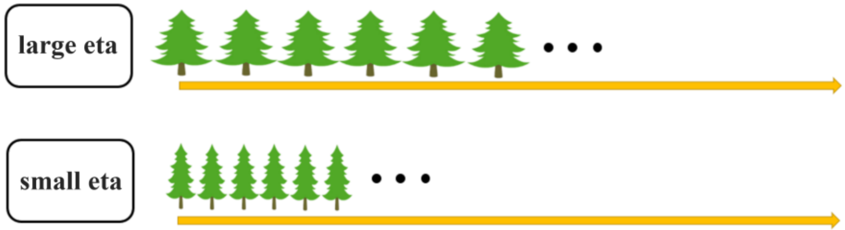
Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

XGBoost parameter - - ensemble

- **learning_rate (eta): step size shrinkage used in tree grow**

$$\hat{y}_i^{(k+1)} = \hat{y}_i^{(k)} + \eta f_{k+1}(x_i)$$



- **Influence of learning_rate on the XGBoost model ?**

Objective function

- **Logistic regression & SVM - - fixed**

- **Ensemble model - - optional: differentiable & can be optimized**

➢ **RMSE, error, log_loss…**
➢ **only measures model's generalization ability**

- **XGBoost: model performance + computing speed**

Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

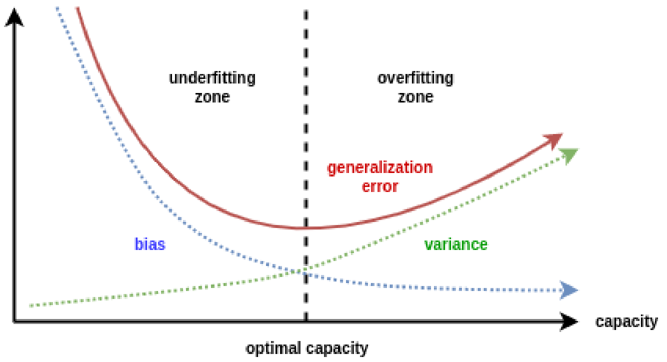
XGBoost objective function

- **Obj = loss function + model complexity**

$$Obj = \boxed{\sum_{i=1}^{m} l(y_i, \hat{y}_i)} + \boxed{\sum_{k=1}^{K} \Omega(f_k)}$$

$$\hat{y}_i^{(t)} = \sum_{k}^{t} f_k(x_i) = \sum_{k}^{t-1} f_k(x_i) + f_t(x_i)$$

$$E(f; D) = bias^2 + var + \epsilon^2$$



XGBoost objective function

- **Solve the objective function**

$$Obj = \boxed{\sum_{i=1}^{m} l(y_i, \hat{y}_i)} + \boxed{\sum_{k=1}^{K} \Omega(f_k)}$$

at *t*-th iteration

$$\boxed{\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)}$$

$$= \sum_{i=1}^{m} \boxed{l(y_i^t, \hat{y}_i^{(t-1)} + f_t(x_i))} + \sum_{k=1}^{t-1} \Omega(f_k) + \Omega(f_t)$$

$$F(\hat{y}_i^{(t-1)} + f_t(x_i))$$

XGBoost objective function

- **Taylor Expansion**

$$f(x) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \frac{f'''(c)}{3!}(x-c)^3 + \ldots$$

assume *c* close to *x*

$$f(x+x-c) \approx \frac{f(c)}{0!} + \frac{f'(c)}{1!}(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \frac{f'''(c)}{3!}(x-c)^3 + \ldots$$

$$\approx \frac{f(c)}{0!} + \frac{f'(c)}{1!}(x-c) + \frac{f''(c)}{2!}(x-c)^2$$

$$\approx f(c) + f'(c)(x-c) + \frac{f''(c)}{2}(x-c)^2$$

$x_1 = x$, $x_2 = x-c$

$$f(x_1 + x_2) \approx f(x_1) + f'(x_1) * x_2 + \frac{f''(x_1)}{2} * x_2^2$$

Department of Mechanics and Maritime Sciences
Division of Marine Technology

XGBoost objective function

- **Solve the objective function**

$x_1 = x$, $x_2 = x\text{-}c$

$$f(x_1 + x_2) \approx f(x_1) \quad + \quad x_2 * f'(x_1) \quad + \quad \frac{1}{2}(x_2)^2 * f''(x_1)$$

$$F(\hat{y}_i^{(t-1)} + f_t(x_i)) \approx F(\hat{y}_i^{(t-1)}) \quad + f_t(x_i) * \frac{\partial F(\hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \quad + \frac{1}{2}(f_t(x_i))^2 * \frac{\partial^2 F(\hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}$$

at *t*-th iteration

$$\approx l(y_i^t, \hat{y}_i^{(t-1)}) + f_t(x_i) * \frac{\partial l(y_i^t, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} + \frac{1}{2}(f_t(x_i))^2 * \frac{\partial^2 l(y_i^t, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}$$

$$\approx l(y_i^t, \hat{y}_i^{(t-1)}) + f_t(x_i) * g_i \quad + \frac{1}{2}(f_t(x_i))^2 * h_i$$

XGBoost objective function

- **Solve the objective function**

at *t*-th iteration

$$= \sum_{i=1}^{m}[l(y_i^t, \hat{y}_i^{(t-1)}) + f_t(x_i)g_i + \frac{1}{2}(f_t(x_i))^2 h_i)] \quad + \quad \sum_{k=1}^{t-1}\Omega(f_k) + \Omega(f_t)$$

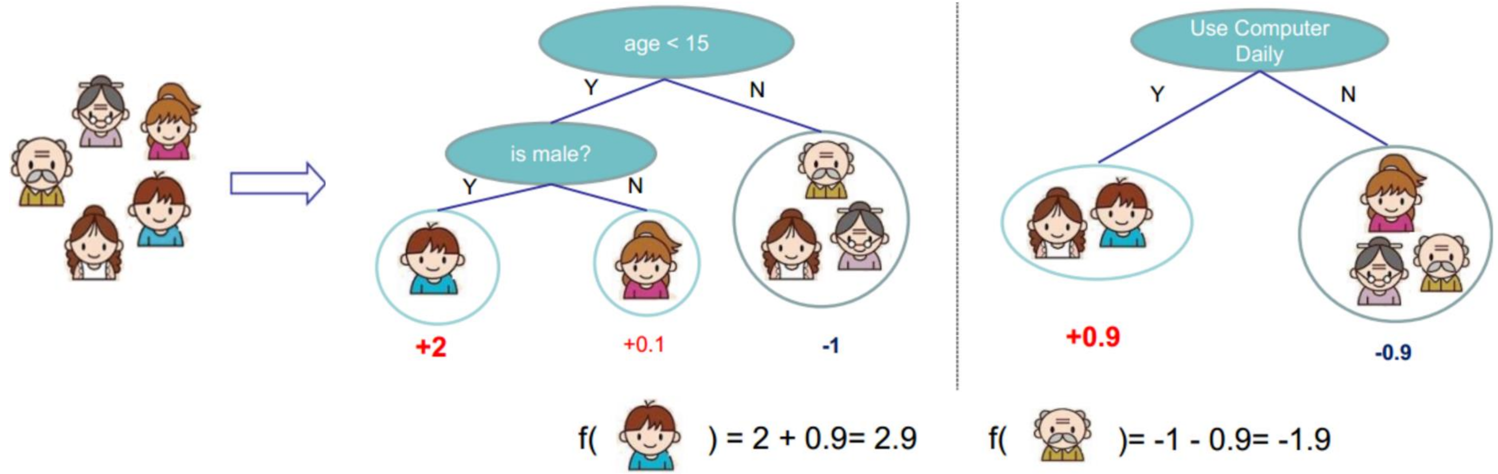
$$Obj = \sum_{i=1}^{m}[f_t(x_i)g_i + \frac{1}{2}(f_t(x_i))^2 h_i)] + \Omega(f_t)$$

Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

XGBoost parameter - - weak evaluator

- **alpha (*L*1 regularization) & lambda (*L*2 regularization)**



XGBoost parameter - - weak evaluator

- **alpha (*L*1 regularization) & lambda (*L*2 regularization)**

$$f_t(x_i) = w_{q(x_i)}$$

$$\Omega(f) = \gamma T + Regularization$$

$$= \gamma T + \frac{1}{2}\alpha|w|$$

$$= \gamma T + \frac{1}{2}\alpha\sum_{j=1}^{T}|w_j|$$

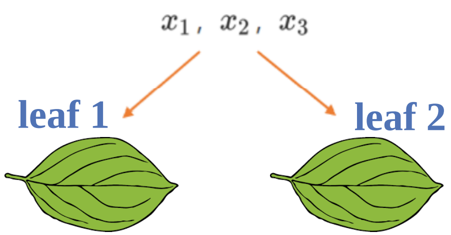
$$= \gamma T + \frac{1}{2}\alpha\sum_{j=1}^{T}|w_j| + \frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2$$

$$= \gamma T + \frac{1}{2}\lambda||w||^2$$

$$= \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2$$

XGBoost tree structures

- $\boldsymbol{\omega}$ **&** $\boldsymbol{T}$



$x_1,\ x_2$ $\qquad\qquad$ x_3

$w_{q(x_1)} * g_1$
$w_{q(x_2)} * g_2$ $\qquad\qquad$ $w_{q(x_3)} * g_3$

$$w_{q(x_1)} = w_{q(x_2)} = w_1 \qquad w_{q(x_3)} = w_2$$

$$\begin{aligned}\sum_{i=1}^{m} w_{q(x_i)} * g_i &= w_{q(x_1)} * g_1 + w_{q(x_2)} * g_2 + w_{q(x_3)} * g_3 \\ &= w_1(g_1 + g_2) + w_2 * g_3 \\ &= \sum_{j=1}^{T} (w_j \sum_{i \in I_j} g_i)\end{aligned}$$

$$\begin{aligned}&\sum_{i=1}^{m} [f_t(x_i) g_i + \frac{1}{2} (f_t(x_i))^2 h_i] + \Omega(f_t) \\ &= \sum_{i=1}^{m} [w_{q(x_i)} g_i + \frac{1}{2} w_{q(x_i)}^2 h_i] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \\ &= \boxed{\sum_{i=1}^{m} w_{q(x_i)} g_i} + \boxed{\sum_{i=1}^{m} \frac{1}{2} w_{q(x_i)}^2 h_i} + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2\end{aligned}$$

Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

XGBoost tree structures

- ω & T

$$\sum_{j=1}^{T}(w_j * \sum_{i \in I_j} g_i) \; + \; \frac{1}{2}\sum_{j=1}^{T}(w_j^2 * \sum_{i \in I_j} h_i) \; + \; \gamma T \; + \; \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

$$= \sum_{j=1}^{T}\left[w_j \sum_{i \in I_j} g_i + \frac{1}{2} w_j^2 (\sum_{i \in I_j} h_i + \lambda)\right] + \gamma T$$

$$G_j = \sum_{i \in I_j} g_i, \quad H_j = \sum_{i \in I_j} h_i$$

$$= \sum_{j=1}^{T}\left[w_j G_j + \frac{1}{2} w_j^2 (H_j + \lambda)\right] + \gamma T$$

$$F^*(w_j) = w_j G_j + \frac{1}{2} w_j^2 (H_j + \lambda)$$

$$\frac{\partial F^*(w_j)}{\partial w_j} = G_j + w_j(H_j + \lambda)$$

$$0 = G_j + w_j(H_j + \lambda)$$

$$w_j = -\frac{G_j}{H_j + \lambda}$$

Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

XGBoost tree structures

- $\boldsymbol{\omega}$ **&** $\boldsymbol{T}$

$$= \sum_{j=1}^{T}\left[-\frac{G_j}{H_j+\lambda} * G_j + \frac{1}{2}\left(-\frac{G_j}{H_j+\lambda}\right)^2 (H_j+\lambda)\right] + \gamma T$$

$$= \sum_{j=1}^{T}\left[-\frac{G_j^2}{H_j+\lambda} + \frac{1}{2} * \frac{G_j^2}{H_j+\lambda}\right] + \gamma T$$

$$= \boxed{-\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda} + \gamma T}$$

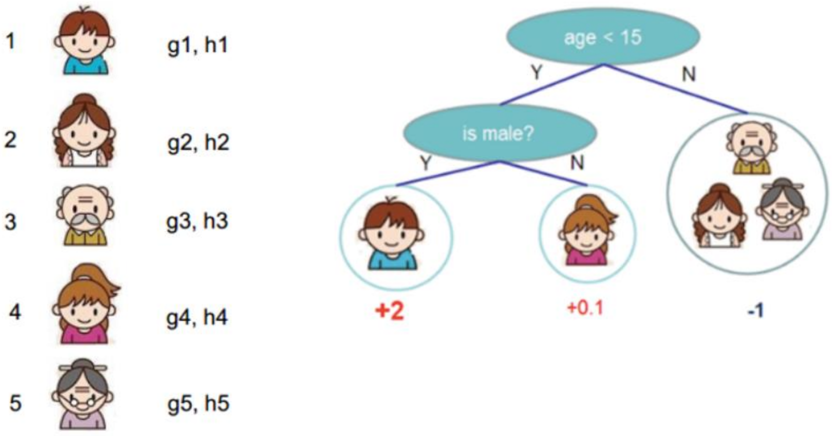
structure score

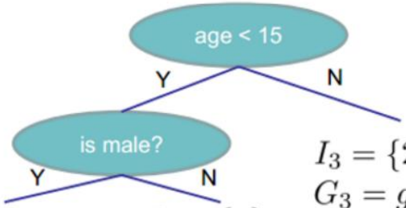
Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

XGBoost tree structure

- **Structure score**





$$I_1 = \{1\} \quad G_1 = g_1 \quad H_1 = h_1$$

$$I_2 = \{4\} \quad G_2 = g_4 \quad H_4 = h_4$$

$$I_3 = \{2, 3, 5\} \quad G_3 = g_2 + g_3 + g_5 \quad H_3 = h_2 + h_3 + h_5$$

$$Obj = -\sum_j \frac{G_j^2}{H_j + \lambda} + 3\gamma$$

$$Obj = -\left(\frac{g_1^2}{h_1 + \lambda} + \frac{g_4^2}{h_4 + \lambda} + \frac{(g_2 + g_3 + g_5)^2}{h_2 + h_3 + h_5 + \lambda}\right) + 3\gamma$$

Department of Mechanics and Maritime Sciences
Division of Marine Technology

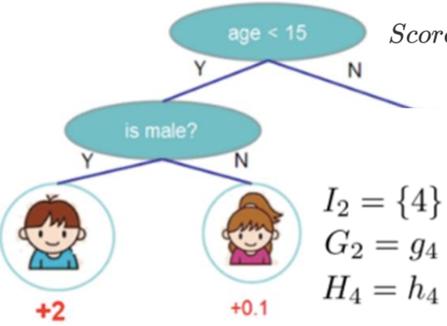
Xiao Lang, 2021-04-30

XGBoost tree structure

- **Greedy learning**

$$I = \{1, 4\}$$
$$G = g_1 + g_4$$
$$H = h_1 + h_4$$
$$Score_{middle} = -\frac{1}{2}\frac{G^2}{H+\lambda} + \gamma$$

$$I_1 = \{1\}$$
$$G_1 = g_1$$
$$H_1 = h_1$$



$$I_2 = \{4\}$$
$$G_2 = g_4$$
$$H_4 = h_4$$

$$Score_{bro} = -\frac{1}{2}\frac{g_1^2}{h_1+\lambda} + \gamma \qquad Score_{sis} = -\frac{1}{2}\frac{g_4^2}{h_4+\lambda} + \gamma$$

$$\begin{aligned}
Gain &= Score_{sis} + Score_{bro} - Score_{middle} \\
&= -\frac{1}{2}\frac{g_4^2}{h_4+\lambda} + \gamma - \frac{1}{2}\frac{g_1^2}{h_1+\lambda} + \gamma - \left(-\frac{1}{2}\frac{G^2}{H+\lambda} + \gamma\right) \\
&= -\frac{1}{2}\frac{g_4^2}{h_4+\lambda} + \gamma - \frac{1}{2}\frac{g_1^2}{h_1+\lambda} + \gamma + \frac{1}{2}\frac{G^2}{H+\lambda} - \gamma \\
&= -\frac{1}{2}\left[\frac{g_4^2}{h_4+\lambda} + \frac{g_1^2}{h_1+\lambda} - \frac{G^2}{H+\lambda}\right] + \gamma \\
&= -\frac{1}{2}\left[\frac{g_4^2}{h_4+\lambda} + \frac{g_1^2}{h_1+\lambda} - \frac{(g_1+g_4)^2}{(h_1+h_4)+\lambda}\right] + \gamma
\end{aligned}$$

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma$$

Department of Mechanics and Maritime Sciences
Division of Marine Technology

Xiao Lang, 2021-04-30

XGBoost parameter - - weak evaluator

- $\boldsymbol{\gamma}$ $$Gain = \frac{1}{2}[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}] - \gamma$$

$$\frac{1}{2}[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}] - \gamma > 0$$

$$\frac{1}{2}[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}] > \gamma$$

Summary

- **Efficiency**
  - ➢ **support parallel processing implementation**
- **Accuracy**
  - ➢ **in-built regularization to reduce overfitting**
  - ➢ **more effective tree pruning**
- **Feasibility**
  - ➢ **customized objective and evaluation**
  - ➢ **tunable parameters**

AVANCEZ

1829

CHALMERS

UNIVERSITY OF TECHNOLOGY