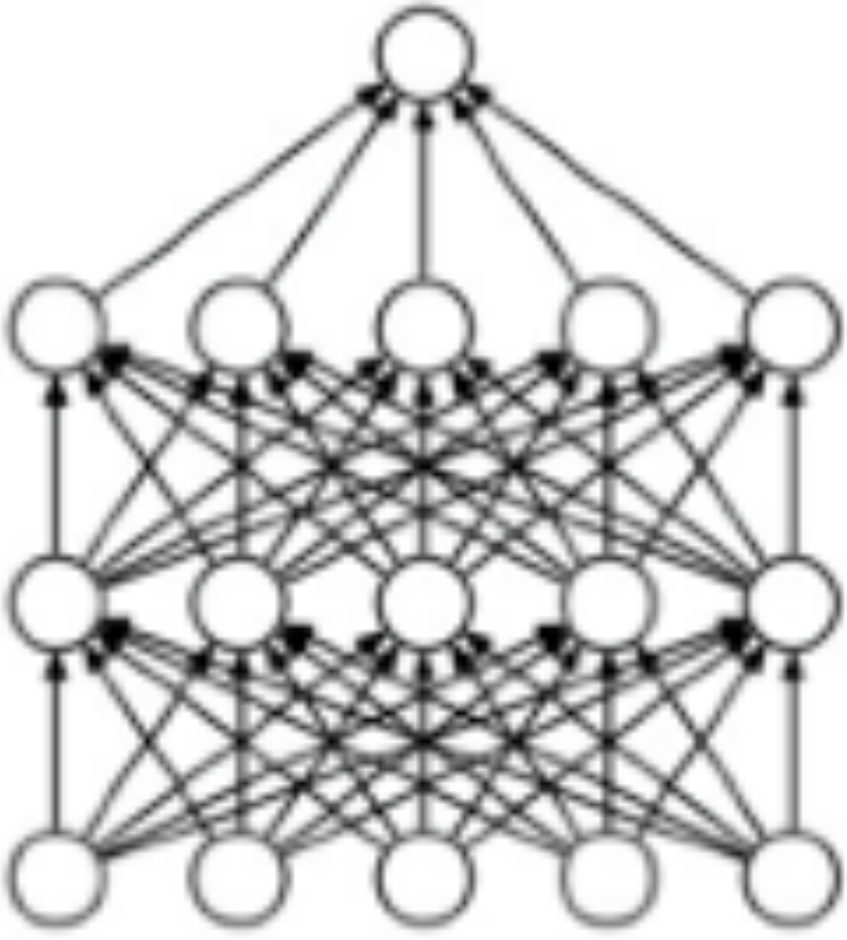

H2O.ai

Dropouts: DART

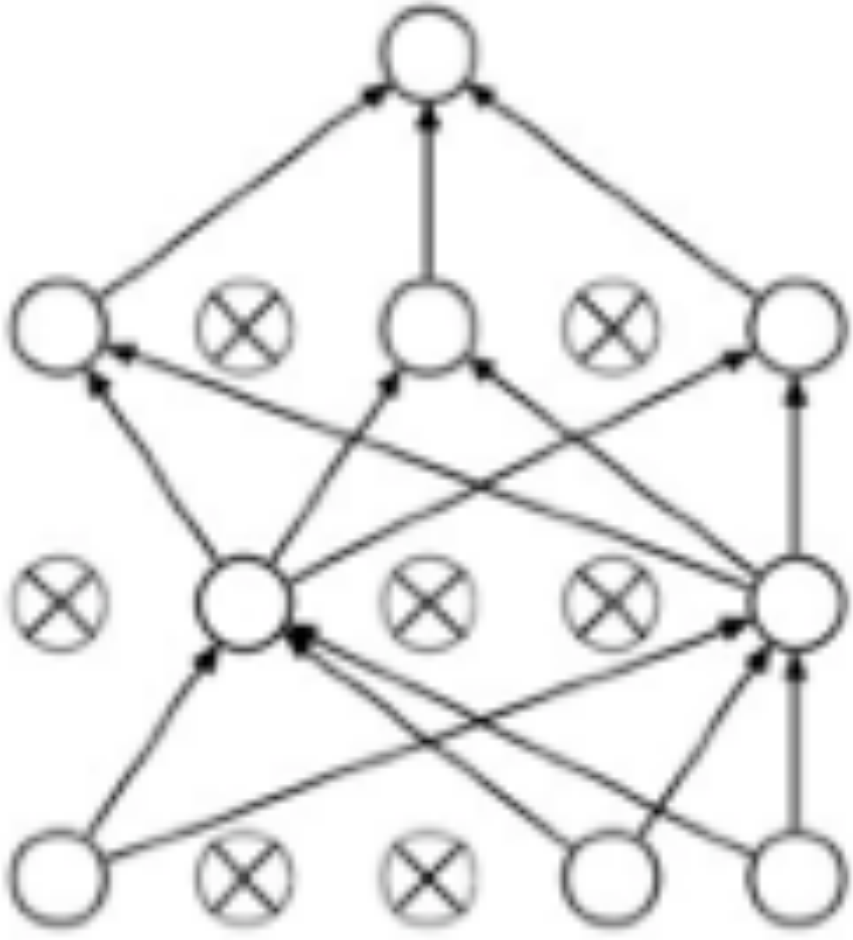
- Idea of dropouts comes from deep neural networks (CNN)

- DART: Dropouts meet Multiple Additive Regression Trees

Rashmi, K. and Gilad, R. (2015). DART: Dropouts meet Multiple Additive Regression Trees



(a) Standard Neural Net



(b) After applying dropout.

Algorithm 1 The DART algorithm

Let N be the total number of trees to be added to the ensemble
$S_1 \leftarrow \{\boldsymbol{x}, -L'_{\boldsymbol{x}}(0)\}$
T_1 be a tree trained on the dataset S_1
$M \leftarrow \{T_1\}$
for $t = 2, \ldots, N$ **do**
    $D \leftarrow$ the subset of M such that $T \in M$ is in D with probability p_{drop}
    if $D = \emptyset$ **then** $D \leftarrow$ a random element from M
    end if
    $\hat{M} \leftarrow M \setminus D$
    $S_t \leftarrow \left\{\boldsymbol{x}, -L'_{\boldsymbol{x}}\left(\hat{M}(\boldsymbol{x})\right)\right\}$
    T_t be a tree trained on the dataset S_t
    $M \leftarrow M \cup \left\{\frac{T_t}{|D|+1}\right\}$
    for $T \in D$ **do**
        Multiply T in M by a factor of $\frac{|D|}{|D|+1}$
    end for
end for
Output M

1) Use only a subset of previous trees for computing gradient for next tree

2) Normalization factors to account for the fact trees were removed

GPU XGBoost

Accelerating the XGBoost algorithm using GPU computing

Artificial Intelligence | Data Mining and Machine Learning

Rory Mitchell, Eibe Frank

April 4, 2017

GPU Accelerated XGBoost

dmlc.ml/2016/12/14/GPU-accelerated-xgboost.html

DMLC Archive

GPU Accelerated XGBoost

Dec 14, 2016 • Rory Mitchell

Update 2016/12/23: Some of our benchmarks were incorrect due to a wrong compiler flag. These have all been updated below.

Decision tree learning and gradient boosting have until recently been the domain of multicore CPUs. Here we showcase a new plugin providing GPU acceleration for the XGBoost library. The plugin provides significant speedups over multicore CPUs for large datasets.

The plugin can be found at: https://github.com/dmlc/xgboost/tree/master/plugin/updater_gpu

Before talking about the GPU plugin we briefly explain the XGBoost algorithm.

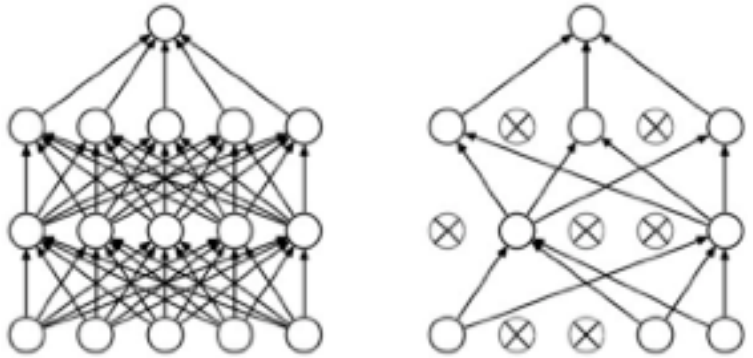
XGBoost for classification and regression

XGBoost is a powerful tool for solving classification and regression problems in a supervised learning setting. It is an implementation of a generalised gradient boosting algorithm designed to offer high-performance, multicore scalability and distributed machine scalability.

The gradient boosting algorithm is an ensemble learning technique that builds many predictive models. Together these smaller models produce much stronger predictions than any single model alone. In particular for gradient boosting, we create these smaller models sequentially, where each new model directly addresses the weaknesses in the previous models.

Dropouts: DART

- Idea of dropouts comes from deep neural networks (CNN)



(a) Standard Neural Net (b) After applying dropout.

- DART: Dropouts meet Multiple Additive Regression Trees

1) Use only a subset of previous trees for computing gradient for next tree

2) Normalization factors to account for the fact trees were removed

Algorithm 1 The DART algorithm

Let N be the total number of trees to be added to the ensemble
$S_1 \leftarrow \{x, -L'_x(0)\}$
T_1 be a tree trained on the dataset S_1
$M \leftarrow \{T_1\}$
for $t = 2, \ldots, N$ **do**
  $D \leftarrow$ the subset of M such that $T \in M$ is in D with probability p_{drop}
  if $D = \emptyset$ **then** $D \leftarrow$ a random element from M
  end if
  $\hat{M} \leftarrow M \setminus D$
  $S_t \leftarrow \left\{x, -L'_x\left(\hat{M}(x)\right)\right\}$
  T_t be a tree trained on the dataset S_t
  $M \leftarrow M \cup \left\{\frac{T_t}{|D|+1}\right\}$
  for $T \in D$ **do**
    Multiply T in M by a factor of $\frac{|D|}{|D|+1}$
  end for
end for
Output M

Rashmi, K. and Gilad, R. (2015). DART: Dropouts meet Multiple Additive Regression Trees