H2O.ai

1. Data parallelization and histogramming (like H2O)

2. Gradient-based one-sided sampling (GOSS):
   ➡ don’t test all the data; prefer data points with large gradients

3. Exclusive Feature Bundling:
   ➡ method of feature engineering to reduce sparsity

4. Leaf-wise growth:

LightGBM

- otherwise includes most of the XGBoost options (e.g. L1/L2 regularization, DART)

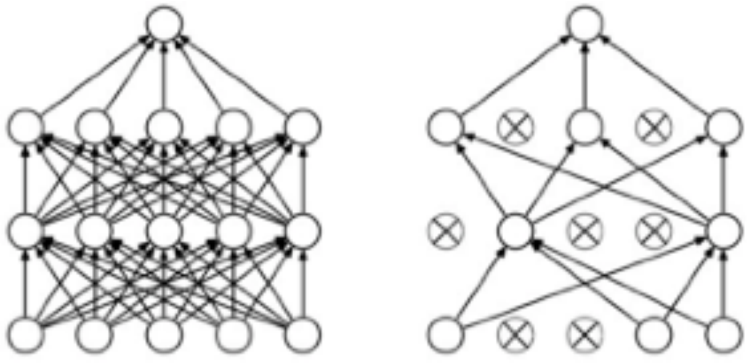
Level-wise tree growth

Leaf-wise tree growth

- released by Microsoft in 2016; attempt to improve speed and memory usage ("Light")

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

LightGBM

- released by Microsoft in 2016; attempt to improve speed and memory usage ("Light")

  1. Data parallelization and histogramming (like H2O)

  2. Gradient-based one-sided sampling (GOSS):
     ➡ don't test all the data; prefer data points with large gradients

  3. Exclusive Feature Bundling:
     ➡ method of feature engineering to reduce sparsity

  4. Leaf-wise growth:

- otherwise includes most of the XGBoost options (e.g. L1/L2 regularization, DART)