XGBoost

Aaryan
CO21BTECH11001

XGBoost (eXtreme Gradient Boost) is an algorithm which follows the principle of gradient boosting. The major difference with GBM is that XGBoost used a more regularized model formalization to control over-fitting, which gives it better performance. It uses boosted trees for regression or classification problems.

Input: $\{x_i, y_i\}_{i=1}^{m}$, a loss function $L(y, f(x))$, max iterations = M, a learning rate α.

1. Initialize the model with a constant value

$$f_0(x) = \underset{\theta}{\operatorname{argmin}} \left(\sum_{i=1}^{m} L(y_i, \theta) \right)$$

   which we can calculate by $\frac{df_0(x)}{d\theta} = 0$

2. Iterate k =1 to M:
   a. Compute the gradients ($g_k(x_i)$) and hessians ($h_k(x_i)$)

   for i =1 **to** m:

$$g_k(x_i) = \left[\frac{\partial L(y_i, f_{k-1}(x_i))}{\partial f_{k-1}(x_i)} \right]$$

$$h_k(x_i) = \left[\frac{\partial^2 L(y_i, f_{k-1}(x_i))}{\partial f_{k-1}(x_i)^2} \right]$$

   end

   b. Fit a base learner (tree) $\phi_k(x)$ with input $\left\{ x_i, -\frac{g_k(x_i)}{h_k(x_i)} \right\}_{i=1}^{m}$ by solving:

$$\phi_k(x) = \underset{\phi}{\operatorname{argmin}} \left(\sum_{i=1}^{m} \frac{1}{2} h_k(x_i) \left[-\frac{g_k(x_i)}{h_k(x_i)} - \phi(x_i) \right]^2 \right)$$

c. Update the model:

$$f_k(x) = f_{k-1}(x) + \alpha\phi_k(x)$$

3. Output function: $f_M(x)$.

Questions:

1. XGBoost is a/an:
   (a) Supervised learning algorithm
   (b) Unsupervised learning algorithm
   Ans. (a)

2. XGBoost can be used for:
   (a) Regression
   (b) Classification
   (c) Both of these
   (d) None of these
   Ans. (c)

3. Which of the following is an important assumption in XGBoost algorithm?
   (a) The value of every feature of data should be between 0 and 1.
   (b) Each sample model ($\phi_k(x)$) is a stump.
   (c) Loss function can be approximated by a second order approximation.
   (d) None of the above
   Ans. (c)

4. Why is regularization used in XGBoost?
   Ans. Regularization prevents overfitting of the model by reducing the dependency of hypothesis function on one particular training data.

5. What are the advantages of using XGBoost?
   Ans. It does not need normalized features and work with non-linear data also.

6. What are the disadvantages of using XGBoost?
   Ans. It is sensitive to outliers. It does not perform so well on sparse and unstructured data.