Gradient Boosted Decision Trees

Intro & Regression

Boosting

Boosting – a method for combining outputs of many “**weak**” classifiers or regressors to produce a powerful “**ensemble**”

Variants of boosting: AdaBoost, BrownBoost, LogitBoost, Gradient Boosting

Big Data:

- Large number of examples
- High dimensionality

One can train very complex (and accurate!) machine learning models by two approaches:

1. Start with a complex model from the very beginning and fit its parameters – **neural network**
2. Build a model iteratively, each step requires training of simple models – **boosting**

Regression

Given a training set: $Z=\{(x_1, y_1),\ldots,(x_n, y_n)\}$
x_i- features, y_i- targets (real values)

Goal is to find $f(x)$ using training set, such as

$$\min \sum_{(x,\, y) \in T} (f(x) - y)^2$$

at test set $T=\{(x_1, y_1),\ldots,(x_n, y_n)\}$

How to build $f(x)$?

Gradient Boosted Trees for Regression

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} h_m(\boldsymbol{x})$$

$\boldsymbol{h_m(x)}$ - a decision tree

Algorithm: Gradient Boosted Trees for Regression

Input: training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

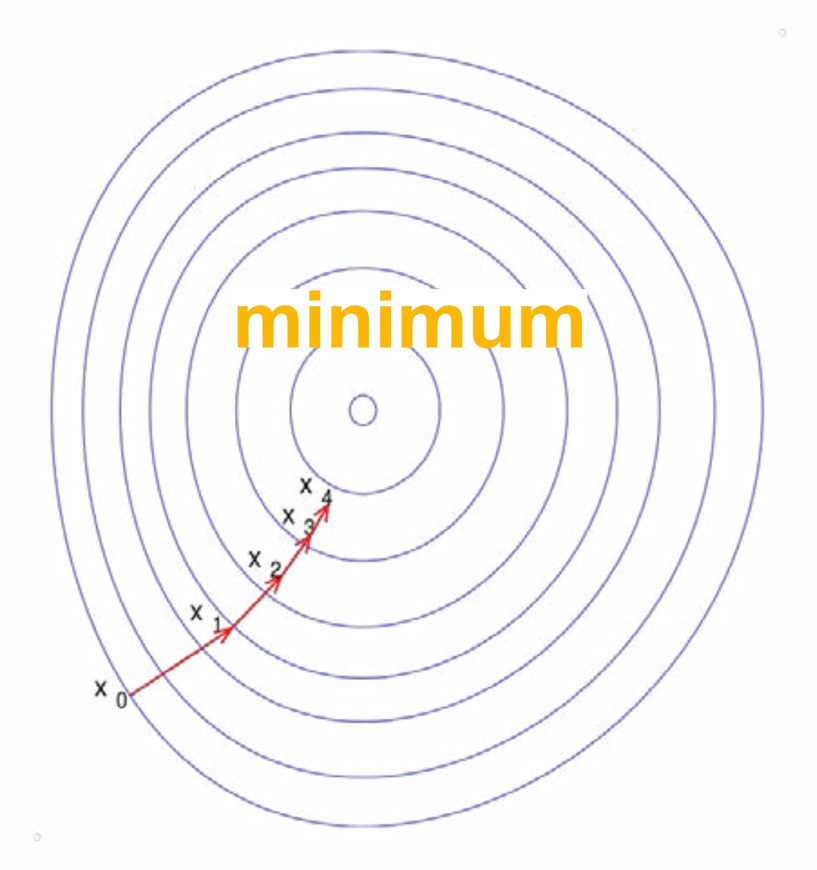
M – number of iterations

1. $f_0(x) = \frac{1}{n}\sum_{i=1}^{n} y_i$
2. For m=1…M:
3. $\hat{y}_i = y_i - f_{m-1}(x_i)$ (residual)
4. Fit a decision tree $h_m(x)$ to the targets $\hat{y}_i$
   (auxiliary training set $\{(x_1, \hat{y}_1), \ldots, (x_n, \hat{y}_n)\}$)
5. $f_m(x) = f_{m-1}(x) + \nu h_m(x)$
6. Return $f_m(x)$

ν - regularization (learningRate), recommended ≤ 0.1

Optimization theory

Function maximization with gradient descent



$$f_m(x) = f_0(x) + vh_1(x) + vh_2(x) + \ldots$$

Boosting – minimization in the functional space

Summary

- **Boosting** – a method for combining outputs of many “**weak**” classifiers or regressors to produce a powerful “**ensemble**”

- **Gradient Boosting** – a gradient descent minimization of the target function in the functional space

- **Gradient Boosting with Decision Trees** – the best algorithm for general purpose classification or regression