

Gradient boosting

Sunday, March 18, 2018

9:37 AM

Gradient boosting = gradient descent + boosting

Recall that AdaBoost

- fits an additive model (ensemble) of the form

$$\sum_{l=1}^L \alpha_l h_l(x)$$

in a forward stagewise manner

- in each stage, a new classifier or regressor is introduced to compensate for the weaknesses of previous classifiers or regressors.
- in AdaBoost, weaknesses are identified by weighting data points
- in gradient boosting, weaknesses of previous classifiers/regressors are identified by gradients.
- in both AdaBoost and gradient descent, weights on data points or gradients guide improvement of the ensemble model.

History

- 1997: Freund and Schapire develop AdaBoost

1999: Breiman reformulates AdaBoost as gradient descent with an exponential loss function

2001: Friedman generalizes AdaBoost with a more general loss function called Gradient Boosting.

Gradient boosting with regression

$$\mathcal{D} = \{(x^{(i)}, y^{(i)}) \mid 1 \leq i \leq m; x^{(i)} \in \mathbb{R}^d; y^{(i)} \in \mathbb{R}\}$$

Suppose you have built a model

$$h_1(x) = \theta_1^T x$$

to minimize sum squared error.

Let's say h_1 is good, but not as good as we would like it to be. Could we build a better model? We will use the ensemble idea and build a new model of the form

$$\underbrace{h_1(x)}_{\text{previous model}} + \underbrace{h_2(x)}_{\text{new model}} \quad \text{an additive ensemble}$$