GBM Tutorial Notes

Creator: YueLiang (Grader)
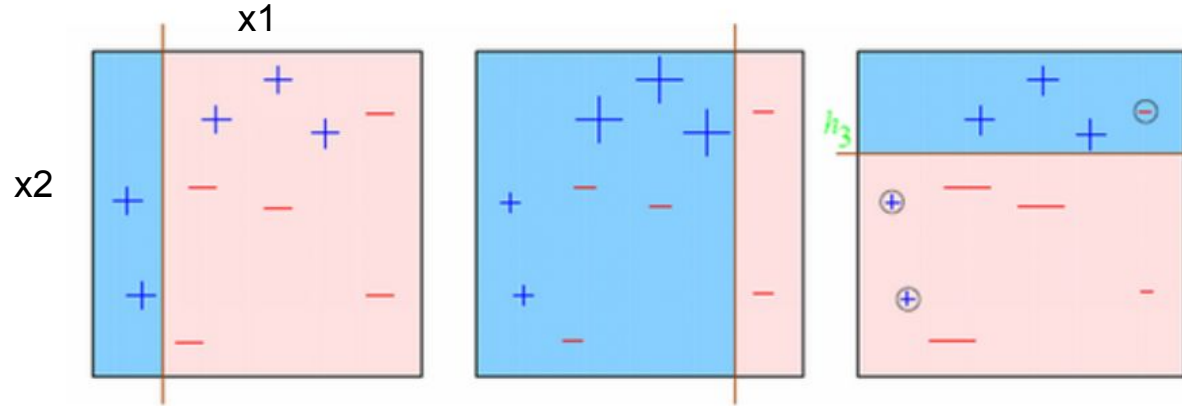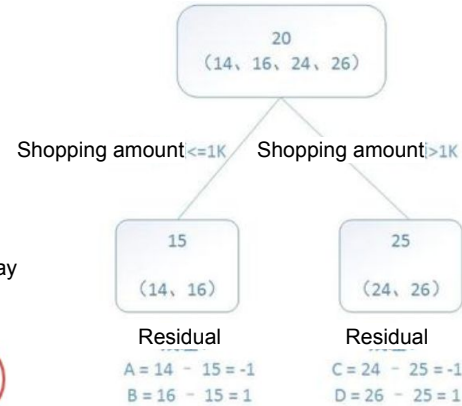
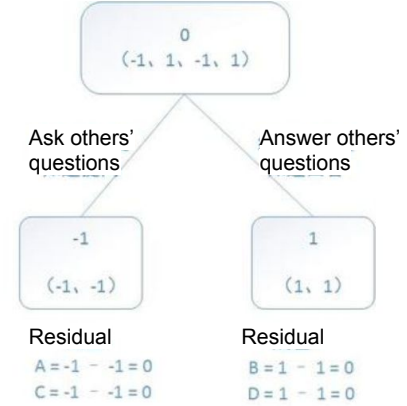| Name	Loss	Derivative	f^*	Algorithm
Squared error	$\frac{1}{2}(y_i - f(\mathbf{x}_i))^2$	$y_i - f(\mathbf{x}_i)$	$\mathbb{E}[y|\mathbf{x}_i]$	L2Boosting
Absolute error	$|y_i - f(\mathbf{x}_i)|$	$\text{sgn}(y_i - f(\mathbf{x}_i))$	$\text{median}(y|\mathbf{x}_i)$	Gradient boosting
Exponential loss	$\exp(-\tilde{y}_i f(\mathbf{x}_i))$	$-\tilde{y}_i \exp(-\tilde{y}_i f(\mathbf{x}_i))$	$\frac{1}{2}\log\frac{\pi_i}{1-\pi_i}$	AdaBoost
Logloss	$\log(1 + e^{-\tilde{y}_i f_i})$	$y_i - \pi_i$	$\frac{1}{2}\log\frac{\pi_i}{1-\pi_i}$	LogitBoost

Tree 1 (left):

20
(14,16,24,26)

Shopping amount<=1K   Shopping amount>1K

15
(14,16)

25
(24,26)

Web-surfing<=1.1h Time   Web-surfing>1.1h Time   Surfing All day   Do not surf All day

14
(14)

16
(16)

24
(24)

26
(26)

http://blog.csdn.net/w2897102
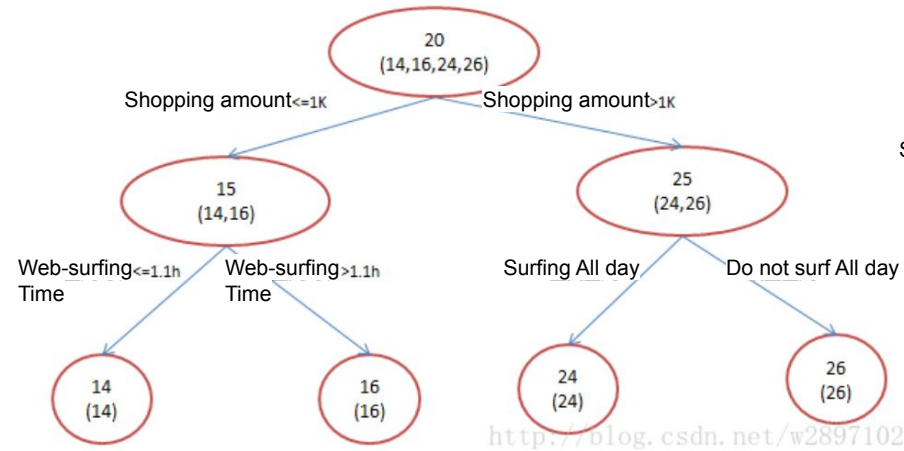
Predict the four people's ages

Tree 2 (middle):

20
(14、16、24、26)

Shopping amount<=1K   Shopping amount>1K

15
(14、16)

25
(24、26)

Residual

A = 14 − 15 = -1
B = 16 − 15 = 1

Residual

C = 24 − 25 = -1
D = 26 − 25 = 1

Tree 3 (right):

0
(-1、1、-1、1)

Ask others' questions   Answer others' questions

-1
(-1、-1)

1
(1、1)

Residual

A = -1 − -1 = 0
C = -1 − -1 = 0

Residual

B = 1 − 1 = 0
D = 1 − 1 = 0

Residual -> Gradient

The **shrinkage** defines the steps taken in the gradient descent of boosting. An important method of regularization. If you take a very small shrinkage you should increase the **interactions** to come to optimum.

Number of trees: The total number of trees to fit. GBMs often require many trees; however, unlike random forests GBMs can overfit so the goal is to find the optimal number of trees that minimize the loss function of interest with cross validation.