

STA 4010, 2018 Fall, CUHKSZ
Assignment 2

- The due is on Sunday, October 14.
 - For programming language, please use R code, along with comments as clear as possible. Points could be deducted even if your codes are correct, but hard to understand.
1. Implement the following algorithm and apply it to the problem in Assignment 1.

Algorithm 10.4 *Gradient Boosting for K-class Classification.*

1. Initialize $f_{k0}(x) = 0$, $k = 1, 2, \dots, K$.

2. For $m=1$ to M :

(a) Set

$$p_k(x) = \frac{e^{f_k(x)}}{\sum_{\ell=1}^K e^{f_\ell(x)}}, \quad k = 1, 2, \dots, K.$$

(b) For $k = 1$ to K :

- i. Compute $r_{ikm} = y_{ik} - p_k(x_i)$, $i = 1, 2, \dots, N$.
- ii. Fit a regression tree to the targets r_{ikm} , $i = 1, 2, \dots, N$, giving terminal regions R_{jkm} , $j = 1, 2, \dots, J_m$.
- iii. Compute

$$\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}} r_{ikm}}{\sum_{x_i \in R_{jkm}} |r_{ikm}|(1 - |r_{ikm}|)}, \quad j = 1, 2, \dots, J_m.$$

iv. Update $f_{km}(x) = f_{k,m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jkm} I(x \in R_{jkm})$.

3. Output $\hat{f}_k(x) = f_{kM}(x)$, $k = 1, 2, \dots, K$.
