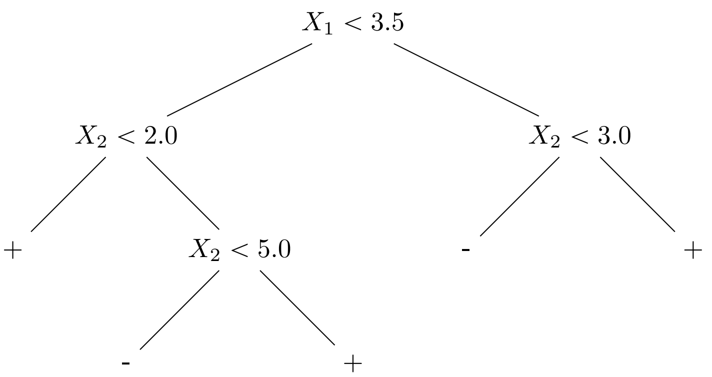
CS680, Spring 2020, Assignment 6

Zhijie Wang, [redacted]

June 23, 2020

Exercise 1

1. We should pick $t \quad 2.0$.
2. We should pick $t \quad 5.0$.
3. I will choose $\mathbf{X}_2$ with $t \quad 5.0$. Because in this case, Gini Index is smaller.



4.

5. Yes. In the first level, $X_1 < 3.5$ is true, will go to left subtree. In the second level, $X_2 < 2.0$ is false, will go to right subtree. In the third level, $X_2 < 5.0$ is false, will go to right child, which is label '+'.

Exercise 2

1. *Proof.* By induction on t. When $t \quad 1$, $p_i^1 \quad \frac{w_i^1}{\sum_{j=1}^n w_j^1}$, $\tilde{p}_i^1 \quad \frac{\tilde{w}_i^1}{\sum_{j=1}^n \tilde{w}_j^1}$, where w^1 and $\tilde{w}^1$ are obviously same.

   When $t \quad m$, assume $p^m \quad \tilde{p}^m$, then, $\frac{w_i^m}{\sum_{j=1}^n w_j^m} \quad \frac{\tilde{w}_i^m}{\sum_{j=1}^n \tilde{w}_j^m}$.

   Now we prove it on $t \quad m+1$.

   When $t \quad m+1$, $p^{m+1} \quad \frac{w_i^{m+1}}{\sum_{j=1}^n w_j^{m+1}} \quad \frac{w_i^m \exp(-y_i \beta_m h_m(\mathbf{x}_i))}{\sum_{j=1}^n w_j^m \exp(-y_j \beta_m h_m(\mathbf{x}_j))} \quad \frac{w_i^m (\frac{1 \quad \epsilon_m}{\epsilon_m}) \quad ^{\frac{1}{2} y_i h_m(\mathbf{x}_i)}}{\sum_{j=1}^n w_j^m \frac{1 \quad \epsilon_m}{\epsilon_m}) \quad ^{\frac{1}{2} y_j h_m(\mathbf{x}_j)}}$.

   Now, $\tilde{p}^{m+1} \quad \frac{\tilde{w}_i^{m+1}}{\sum_{j=1}^n \tilde{w}_j^{m+1}} \quad \frac{\tilde{w}_i^m \tilde{\beta}_m^{1 \quad |\tilde{h}_m(\mathbf{x}_i) \quad \tilde{y}_i|}}{\sum_{j=1}^n \tilde{w}_j^m \tilde{\beta}_m^{1 \quad |\tilde{h}_m(\mathbf{x}_j) \quad \tilde{y}_j|}} \quad \frac{\tilde{w}_i^m (\frac{\epsilon_m}{1 \quad \epsilon_m})^{1 \quad |\tilde{h}_m(\mathbf{x}_i) \quad \tilde{y}_i|}}{\sum_{j=1}^n \tilde{w}_i^m (\frac{\epsilon_m}{1 \quad \epsilon_m})^{1 \quad |\tilde{h}_m(\mathbf{x}_j) \quad \tilde{y}_j|}}$

   $$\tilde{p}^{m+1} \quad \frac{\tilde{w}_i^m (\frac{1 \quad \epsilon_m}{\epsilon_m})^{|\tilde{h}_m(\mathbf{x}_i) \quad \tilde{y}_i| \quad 1}}{\sum_{j=1}^n \tilde{w}_i^m (\frac{1 \quad \epsilon_m}{\epsilon_m})^{|\tilde{h}_m(\mathbf{x}_j) \quad \tilde{y}_j| \quad 1}} \quad \frac{\tilde{w}_i^m (\frac{1 \quad \epsilon_m}{\epsilon_m})^{(\quad \frac{1}{2} y_i h_m(\mathbf{x}_i) \quad \frac{1}{2})}}{\sum_{j=1}^n \tilde{w}_i^m (\frac{1 \quad \epsilon_m}{\epsilon_m})^{(\quad \frac{1}{2} y_j h_m(\mathbf{x}_j) \quad \frac{1}{2})}} \quad \frac{\tilde{w}_i^m (\frac{1 \quad \epsilon_m}{\epsilon_m}) \quad ^{\frac{1}{2} y_i h_m(\mathbf{x}_i)}}{\sum_{j=1}^n \tilde{w}_i^m (\frac{1 \quad \epsilon_m}{\epsilon_m}) \quad ^{\frac{1}{2} y_j h_m(\mathbf{x}_j)}}$$

   Therefore, $\tilde{p}^{m+1}$ and p^m are indeed equal because $\frac{w_i^m}{\sum_{j=1}^n w_j^m} \quad \frac{\tilde{w}_i^m}{\sum_{j=1}^n \tilde{w}_j^m}$.

   Thus, the updates are exactly are equivalent to what we learned in class.

2. *Proof.*

$$\mathbb{E}\left[\exp(-yH)|\mathbf{X} \quad \mathbf{x}\right] \quad \exp(H)\cdot P(y \quad -1|\mathbf{X} \quad \mathbf{x}) + \exp(-H)\cdot P(y \quad 1|\mathbf{X} \quad \mathbf{x}) \tag{1}$$

To minimize (1), let $\dfrac{\partial \mathbb{E}}{\partial H} \quad 0$.

$$\begin{aligned} \frac{\partial \mathbb{E}\left[\exp(-yH)|\mathbf{X} \quad \mathbf{x}\right]}{\partial H} &\quad \exp(H)\cdot P(y \quad -1|\mathbf{X} \quad \mathbf{x}) - \exp(-H)\cdot P(y \quad 1|\mathbf{X} \quad \mathbf{x}) \quad 0 \\ H &\quad \frac{1}{2}\log\frac{P(y \quad 1|\mathbf{X} \quad \mathbf{x})}{P(y \quad -1|\mathbf{X} \quad \mathbf{x})} \end{aligned} \tag{2}$$

Therefore, the exponential loss is proportional to the log odds ratio.

3. When $\bar{h}_t(\mathbf{x}) \quad -h_t(\mathbf{x})$, $\bar{\epsilon}_t \quad \sum_{i=1}^{n} p_i^t \cdot [[\bar{h}_t(\mathbf{x}_i) \neq y_i]] \quad 1-\epsilon_t$, $\bar{\beta}_t \quad \dfrac{\bar{\epsilon}_t}{1-\bar{\epsilon}_t} \quad -\beta_t$.

   Since $\exp(-y_i\bar{\beta}_t \cdot -h_t(\mathbf{x}_i)) \quad \exp(-y_i\beta_t h_t(\mathbf{x}_i))$, therefore the update in w is the same.

4. *Proof.*

$$\begin{aligned} \hat{\mathbb{E}}_t \exp[-y\beta h_t(\mathbf{x})] &\quad P(h_t(\mathbf{x}) \quad y)\exp[-y\beta h_t(\mathbf{x})] + P(h_t(\mathbf{x}) \neq y)\exp[-y\beta h_t(\mathbf{x})] \\ &\quad (1-\epsilon)\exp(-\beta) + \epsilon\exp(\beta) \end{aligned} \tag{3}$$

To minimize (3), let let $\dfrac{\partial \hat{\mathbb{E}}_t \exp[-y\beta h_t(\mathbf{x})]}{\partial \beta} \quad 0$.

$$\begin{aligned} \frac{\partial \hat{\mathbb{E}}_t \exp[-y\beta h_t(\mathbf{x})]}{\partial \beta} &\quad \epsilon_t\exp(\beta) - (1-\epsilon_t)\exp(-\beta) \quad 0 \\ \beta &\quad \frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t} \end{aligned} \tag{4}$$

5.

$$\begin{aligned} \epsilon_{t+1}(h_t) &\quad \sum_{i=1}^{n} p_i^{t+1}\cdot[[h_t(\mathbf{x}_i) \neq y_i]] \quad \sum_{i=1}^{n} \frac{w_i^{t+1}}{\sum_{j=1}^{n} w_j^{t+1}} \cdot [[h_t(\mathbf{x}_i)\neq y_i]] \\ &\quad \sum_{i=1}^{n} \frac{w_i^t\exp(-y_i\beta_t h_t(\mathbf{x}_i))}{\sum_{j=1}^{n} w_j^t \exp(-y_j\beta_t h_t(\mathbf{x}_j))}\cdot[[h_t(\mathbf{x}_i)\neq y_i]] \\ &\quad \sum_{y_i \neq h_t(\mathbf{x}_i)} \frac{w_i^t\exp(\beta_t)}{\sum_{y_i\neq h_t(\mathbf{x}_j)} w_j^t\exp(\beta_t) + \sum_{y_i = h_t(\mathbf{x}_j)} w_j^t \exp(-\beta_t)} \\ &\text{now we multiply a } \frac{1}{\sum_{k=1}^{n} w_k^t} \text{ on both numerator and denominator} \\ &\quad \sum_{y_i \neq h_t(\mathbf{x}_i)} \frac{\frac{w_i^t}{\sum_{k \quad 1}^{n} w_k^t}\exp(\beta_t)}{\sum_{y_i\neq h_t(\mathbf{x}_j)} \frac{w_j^t}{\sum_{k \quad 1}^{n} w_k^t}\exp(\beta_t) + \sum_{y_i = h_t(\mathbf{x}_j)} \frac{w_j^t}{\sum_{k \quad 1}^{n} w_k^t}\exp(-\beta_t)} \\ &\quad \sum_{y_i \neq h_t(\mathbf{x}_i)} \frac{p_i^t\exp(\beta_t)}{P(y\neq h_t(\mathbf{x}))\cdot\exp(\beta_t) + P(y \quad h_t(\mathbf{x}))\cdot\exp(-\beta_t)} \\ &\quad \frac{P(y\neq h_t(\mathbf{x}))\cdot\exp(\beta_t)}{P(y\neq h_t(\mathbf{x}))\cdot\exp(\beta_t) + P(y \quad h_t(\mathbf{x}))\cdot\exp(-\beta_t)} \\ &\quad \frac{\epsilon_t\exp(\beta_t)}{\epsilon_t\exp(\beta_t) + (1-\epsilon_t)\exp(-\beta_t)} \\ &\quad \frac{\epsilon_t}{\epsilon_t + (1-\epsilon_t)\exp(-2\cdot\frac{1}{2}\log\frac{1 \quad \epsilon_t}{\epsilon_t})} \\ &\quad \frac{\epsilon_t}{\epsilon_t + (1-\epsilon_t)\frac{\epsilon_t}{(1 \quad \epsilon_t)}} \\ &\quad \frac{1}{2} \end{aligned} \tag{5}$$