H/W 5

Anurag Dhaipule

4.1. $\ell(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2 = \frac{1}{2}(f(x) - y)^2$

Given $$g_m = \left(\frac{\partial}{\partial f(x_i)} \sum_{i=1}^{n} \ell(y_i, f(x_i)) \Big|_{f(x_i) = f_{m-1}(x_i)} \right)_{i=1}^{n}$$

$$g_m = \left(\frac{\partial}{\partial f(x_i)} \sum_{i=1}^{n} \frac{1}{2}(\overset{f(x_i)}{\cancel{\hat{y}}} - y)^2 \Big|_{f(x_i) = f_{m-1}(x_i)} \right)_{i=1}^{n}$$

$$g_m = \left(\sum_{i=1}^{n} (f(x_i) - y) \Big|_{f(x_i) = f_{m-1}(x_i)} \right)_{i=1}^{n}$$

$$g_{m_i} = \cancel{y}\ \cancel{\ell(}\ f_{m-1}(x_i) - y_i$$

Given,

$$h_m = \underset{h \in F}{\arg\min} \sum_{i=1}^{n} ((-g_m)_i - h(x_i))^2$$

$$= \underset{h \in F}{\arg\min} \sum_{i=1}^{n} [y_i - f_{m-1}(x_i) - h(x_i)]^2$$

2. $l(m) = \ln(1+e^{-m}) = \ln(1+e^{-yf(x)}) = l(y, f(x))$

$$g_m = \left(\frac{\partial}{\partial f(x_i)} \sum_{i=1}^{n} l(y_i, f(x_i)) \Big|_{f(x) = f_{m-1}(x_i)} \right)_{i=1}^{n}$$

$$g_m = \left(\sum_{i=1}^{n} \frac{\partial}{\partial f(x_i)} \ln(1+e^{-y_i f(x_i)}) \Big|_{f(x_i) = f_{m-1}(x_i)} \right)_{i=1}^{n}$$

$$g_m = \left(\sum_{i=1}^{n} \frac{1}{1+e^{-y_i f(x_i)}} \cdot e^{-y_i f(x_i)} \cdot y_i \Big|_{f(x_i) = f_{m-1}(x_i)} \right)_{i=1}^{n}$$

$$g_{m_i} = \frac{-e^{-y_i f_{m-1}(x_i)}}{1+e^{-y_i f_{m-1}(x_i)}} \cdot y_i$$

Given,

$$h_m = \underset{h \in F}{\arg\min} \sum_{i=1}^{n} \left((-g_m)_i - h(x_i) \right)^2$$

$$h_m = \underset{h \in F}{\arg\min} \sum_{i=1}^{n} \left[\frac{e^{-y_i f_{m-1}(x_i)}}{1+e^{-y_i f_{m-1}(x_i)}} \cdot y_i - h(x_i) \right]^2$$

5) 1. $E_y[l(yf(x)|x]$

Given x, y can either be 1 or -1, so, the expectation is given by

$$= \pi(x)\, l(f(x))\Big|_{y=1} + (1-\pi(x))\, l(-f(x))\Big|_{y=-1}$$

Where $\pi(x) = P(y=1|x)$

2. To determine Bayes prediction, we differentiate w.r.t $f(x)$ because we want minimum expectation of loss.

$$\frac{\partial E_y[l(yf(x)|x]}{\partial f(x)} = \frac{\partial}{\partial f(x)}\left[\pi(x)\, l(f(x))|_{y=1} + (1-\pi(x))\, l(-f(x))|_{y=-1}\right]$$

$$\cancel{l(f(x))}\; l(y, f(x)) = e^{-yf(x)} \quad \textcircled{2}$$

$$l(f(x)) = e^{-f(x)}\Big|_{y=1},\; e^{f(x)}\Big|_{y=-1}$$

$$\frac{\partial E_y[l(yf(x))\,x]}{\partial f(x)} = \frac{\partial}{\partial f(x)}\left[\pi(x)\, e^{-f(x)} + (1-\pi(x))\, e^{+f(x)}\right]$$

$$= \pi(x)\, e^{-f(x)} - (1-\pi(x))\, e^{+f(x)} = 0$$

$$= \pi(x) e^{-f(x)} = (1-\pi(x)) e^{f(x)} \quad -① $$

$$\Rightarrow \frac{\pi(x)}{1-\pi(x)} = e^{2f(x)}$$

$$f(x) = \frac{1}{2} \ln \frac{\pi(x)}{1-\pi(x)} = f^*(x)$$

This $f(x)$ gives minimum loss and hence it is bayes prediction function.

From ①, we can take $\pi(x)$ terms on one side and get

$$\pi(x) = \frac{1}{1+e^{-2f^*(x)}}$$

3. $l(y, f(x)) = \ln(1+e^{-yf(x)})$

$$l(f(x)) = \ln(1+e^{-f(x)}) \Big|_{y=1}, \quad \ln(1+e^{f(x)})$$

From eq ②,

$$\frac{\partial E_y[l(yf(x)|x)]}{\partial f(x)} = \frac{\partial}{\partial f(x)} \left[\pi(x) \ln(1+e^{-f(x)}) + (1-\pi(x)) \ln(1+e^{f(x)})\right]$$

$$\Rightarrow \frac{\pi(x)}{1+e^{-f(x)}} \cdot e^{-f(x)} + \frac{(1-\pi(x))}{1+e^{f(x)}} \cdot e^{f(x)} = 0$$

$$\frac{\pi(x)}{1+e^{-f(x)}} \cdot e^{-f(x)} = \frac{1-\pi(x)}{1+e^{f(x)}} \cdot e^{f(x)}$$

$$\frac{\pi(x)}{1-\pi(x)} = e^{f(x)} \cdot \frac{[e^{f(x)}+1]}{e^{f(x)}+1} \quad \leftarrow \mathrm{I}$$

$$\Rightarrow f(x) = \ln \frac{\pi(x)}{1-\pi(x)} = f^*(x)$$

From I, taking $\pi(x)$ terms on one side, we get.

$$\pi(x) = \frac{1}{1+e^{-f^*(x)}}$$

6\. 1. $1(G(x_i) \neq y_i) < \exp(-y\, g(x))$

(i) When prediction is wrong, $G(x_i) \neq y_i$

$$y g(x) = \cancel{\ \ } -1 \qquad \because g(x) = \{-1, +1\}$$

$$\Rightarrow \exp^{(-(-1))} = e > 1$$

(ii) When prediction is correct.

$$0 < e^{-1}$$

$$\Rightarrow 1(G(x_i) \neq y_i) < \exp(-y\, g(x))$$

2\. $L(G, D) = \frac{1}{n} \sum_{i=1}^{n} 1(G(x_i) \neq y_i)$

$$Z_T = \frac{1}{n} \sum_{i=1}^{n} \exp(-y_i f_t(x_i))$$

~~Since~~ from sub part 1,

$$L(G, D) < Z_T$$

7. 1. we have,

$$(\alpha_t, G_t) = \underset{\alpha, G}{\arg\min} \sum_{i=1}^{n} L\left(y_i, f_{t-1}(x_i) + \alpha G(x_i)\right)$$

$$L(y, f(x)) = \exp(-y f(x)).$$

$$\Rightarrow (\alpha_t, G_t) = \underset{\alpha, G}{\arg\min} \sum_{i=1}^{n} \exp\left(-y_i\left[f_{t-1}(x_i) + \alpha G(x_i)\right]\right)$$

$$= \underset{\alpha, G}{\arg\min} \sum_{i=1}^{n} \exp(-y f_{t-1}(x_i) \cdot \exp(-y_i \alpha G(x_i))$$

$$= \underset{\alpha, G}{\arg\min} \sum_{i=1}^{n} w_i^t \exp(-\alpha y_i G(x_i))$$

2. For fixed value of α,
Based on correct or wrong prediction

$$\exp(-\alpha y_i G(x_i)) = \begin{cases} \exp(-\alpha) & \rightarrow \text{correct prediction} \\ \exp(\alpha) & \rightarrow \text{wrong prediction} \end{cases}$$

$$\underset{G}{\arg\min} \sum_{i=1}^{n} w_i^t \exp(-\alpha y_i G(x_i)) = \underset{G}{\arg\min} \sum_{i=1}^{n} w_i^t 1(G(x_i) \neq y_i)$$

Notice that minimizes over both terms doesn't make any difference, hence they should same. As you add up sum if $(G(x_i) \neq y_i)$.