

Let $\hat{p}_{mk(m)} = \max_k \hat{p}_{km}$, so that y_i is classified as $k(m)$, i.e. $\hat{y}_i = k(m)$ in region R_m .

We have:

$$\begin{aligned}
 Q_m(T) &= \frac{1}{N_m} \sum_{\{i: x_i \in R_m\}} I(y_i \neq \hat{y}_i) \\
 &= \frac{1}{N_m} \sum_{\{i: x_i \in R_m\}} \left(1 - I(y_i = \underbrace{\hat{y}_i}_{k(m)}) \right) \\
 &= \frac{1}{N_m} \left(\underbrace{\sum_{\{i: x_i \in R_m\}} 1}_{= N_m} - \sum_{\{i: x_i \in R_m\}} I(y_i = k(m)) \right) \\
 &= 1 - \underbrace{\frac{1}{N_m} \sum_{\{i: x_i \in R_m\}} I(y_i = k(m))}_{\hat{p}_{mk(m)}} \\
 &= 1 - \hat{p}_{mk(m)}
 \end{aligned}$$

$$\theta = (\beta_0, \beta_m, \alpha_{jm}, m=1, \dots, M, j=1, \dots, p)$$

$$R_0(\theta) = \sum_{i=1}^N (y_i - f(x_i))^2 + \lambda \sum_{m=1}^M \sum_{j=1}^p \alpha_{jm}^2 + \lambda \sum_{m=1}^M \beta_m^2$$

$$\rightarrow \frac{\partial R_0(\theta)}{\partial \beta_m} = -2 \sum_{i=1}^N (y_i - f(x_i)) \frac{\partial f(x_i)}{\partial \beta_m} + 2\lambda \beta_m$$

with

$$\begin{aligned} \frac{\partial f(x_i)}{\partial \beta_m} &= \frac{\partial}{\partial \beta_m} g \left(\underbrace{\beta_0 + \sum_{n=1}^M \beta_n \sigma \left(\sum_{j=1}^p \alpha_{jn} x_{ij} \right)}_{T_i} \right) \\ &= g'(T_i) \cdot \frac{\partial T_i}{\partial \beta_m} \\ &= g'(T_i) \cdot \underbrace{\sigma \left(\sum_{j=1}^p \alpha_{jm} x_{ij} \right)}_{= z_{im}} = g'(T_i) \cdot z_{im} \end{aligned}$$

Further, we have:

$$\frac{\partial R_0(\theta)}{\partial \alpha_{jm}} = -2 \sum_{i=1}^N (y_i - f(x_i)) \cdot \frac{\partial f(x_i)}{\partial \alpha_{jm}} + 2\lambda \alpha_{jm}$$

with

$$\begin{aligned} \frac{\partial f(x_i)}{\partial \alpha_{jm}} &= \frac{\partial}{\partial \alpha_{jm}} g \left(\underbrace{\beta_0 + \sum_{n=1}^M \beta_n \sigma \left(\sum_{j=1}^p \alpha_{jn} x_{ij} \right)}_{= T_i} \right) \\ &= g'(T_i) \cdot \frac{\partial T_i}{\partial \alpha_{jm}} \\ &= g'(T_i) \beta_m \sigma' \left(\sum_{j=1}^p \alpha_{jm} x_{ij} \right) \cdot x_{ij} \end{aligned}$$