Weighted update PPR (1)

$$R(\theta) \approx \sum_{i=1}^{N} a_i (b_i - w^T x_i)^2 = (b - Xw)^T A (b - Xw)$$

$$A = \text{diag}(a)$$

$$= w^T X^T A X w - 2 w^T X^T A b + b^T A b$$

$$\frac{\partial R(\theta)}{\partial w} \approx 2 X^T A X w - 2 X A b \doteq 0$$

$$\Rightarrow w = (X^T A X)^{-1} X^T A b$$

Back projection for Cross entropy using softmax

$$R_i(\theta) = \sum_{k=1}^{K} -y_{i,K} \log\left(f_k(x_i)\right) \; ; \; \boxed{f_k(x_i) = \frac{\exp\{\beta_k^T z_i\}}{\sum_{j=1}^{K} \exp\{\beta_j^T z_i\}} \qquad z_{m,i} = \sigma(\alpha_m^T x_i)}$$

insert f ↓

$$R_i(\theta) = -\sum_{k=1}^{K} \left(y_{k,i} \left(\beta_k^T z_i - \log\left(\sum_{j=1}^{K} \exp\{\beta_j z_i\} \right) \right) \right)$$

$$= \left(-\sum_{k=1}^{K} y_{k,i} \beta_k^T z_i \right) + \log\left(\sum_{j=1}^{K} \exp\{\beta_j z_i\} \right) \underbrace{\sum_{k=1}^{K} y_{k,i}}_{=1}$$

$$\frac{\partial R_i(\theta)}{\partial \beta_{k,m}} = -y_{k,i} z_{i,m} + \frac{\exp\{\beta_k z_i\} z_{i,m}}{\underbrace{\sum \exp\{\beta_j z_i\}}_{f_k(x_i)}}$$

$$= \underbrace{-\left(y_{k,i} - f_k(x_i) \right)}_{\delta_{i,k}} z_{i,m}$$

$$\frac{\partial R_i(\theta)}{\partial \alpha_{m,l}} = \underbrace{\left(\sum_{k=1}^{K} \underbrace{-\left(y_{k,i} - f_k(x) \right)}_{\delta_{ik}} \beta_{k,m} \right) \sigma'(\alpha_m^T x_i)}_{S_{i,m}} x_{i,l}$$

BPE

$$S_{im} = \left(\sum_{k=1}^{K} \delta_{i,k} \beta_{k,m} \right) \sigma'(\alpha_m^T x_i)$$