Ex. 10.8 Consider a K-class problem where the targets y_{ik} are coded as 1 if observation i is in class k and zero otherwise. Suppose we have a current model $f_k(x),\ k = 1, \ldots, K$, with $\sum_{k=1}^{K} f_k(x) = 0$ (see (10.21) in Section 10.6). We wish to update the model for observations in a region R in predictor space, by adding constants $f_k(x) + \gamma_k$, with $\gamma_K = 0$.

(a) Write down the multinomial log-likelihood for this problem, and its first and second derivatives.

(b) Using only the diagonal of the Hessian matrix in (1), and starting from $\gamma_k = 0\ \forall k$, show that a one-step approximate Newton update for γ_k is

$$\gamma_k^1 = \frac{\sum_{x_i \in R}(y_{ik} - p_{ik})}{\sum_{x_i \in R} p_{ik}(1 - p_{ik})},\ k = 1, \ldots, K-1, \tag{10.57}$$

where $p_{ik} = \exp(f_k(x_i))/(\sum_{\ell=1}^{K} f_\ell(x_i))$.

(c) We prefer our update to sum to zero, as the current model does. Using symmetry arguments, show that

$$\hat{\gamma}_k = \frac{K-1}{K}(\gamma_k^1 - \frac{1}{K}\sum_{\ell=1}^{K}\gamma_\ell^1),\ k = 1, \ldots, K \tag{10.58}$$

is an appropriate update, where γ_k^1 is defined as in (10.57) for all $k = 1, \ldots, K$.