QUESTION 3

ABHAY SHANKAR K: CS21BTECH11001 & KARTHEEK TAMMANA: CS21BTECH11028

(I)

- Gradient of the error:

$$\nabla E(\mathbf{w}) = \sum_{n=1}^{N} (y_n - t_n)\, \phi_n = \mathbf{\Phi^T}(\mathbf{y} - \mathbf{t})$$

- The Hessian:

$$\nabla\nabla E(\mathbf{w}) = \sum_{n=1}^{N} y_n(1 - y_n)\phi_n\phi_n^T = \mathbf{\Phi^T R \Phi}$$

  where R is the diagonal matrix given by $R_{nn} = y_n(1 - y_n)$

- Update function:

$$\mathbf{w}^{(new)} = \mathbf{w}^{(old)} - \mathbf{H}^{-1}\nabla E(\mathbf{w}) =$$

$$\begin{aligned} \mathbf{w}^{(new)} &= \mathbf{w}^{(old)} - \mathbf{H}^{-1}\nabla E(\mathbf{w}) \\ &= \mathbf{w}^{(old)} - (\mathbf{\Phi^T R \Phi})^{-1}\mathbf{\Phi^T}(\mathbf{y} - \mathbf{t}) \\ &= (\mathbf{\Phi^T R \Phi})^{-1}\left(\mathbf{\Phi^T R \Phi w}^{(old)} - \mathbf{\Phi^T}(\mathbf{y} - \mathbf{t})\right) \\ &= (\mathbf{\Phi^T R \Phi})^{-1}\mathbf{\Phi^T R z} \end{aligned} \tag{1}$$

with

$$\mathbf{z} = \mathbf{\Phi w}^{(old)} - \mathbf{R^{-1}}(\mathbf{y} - \mathbf{t})$$

where $\mathbf{\Phi}$ is the N × M design matrix, whose n'th row is given by ϕ_n^T.

(II) Taking the gradient of the weighted least squares error,

$$\nabla E(\mathbf{w}) = \sum_{n=1}^{N} r_n(\mathbf{w^T}\phi_n - t_n)\phi_n^T = \mathbf{\Phi^T R \Phi w} - \mathbf{\Phi^T R t} \tag{2}$$

with $R_{nn} = r_n$. This is clearly the same form

(III)

Date: October 6, 2023.