4.20 The Hessian matrix for the multi-class logistic regression problem comprises blocks of size M $\times$ M in which block (k, j) (referring to the kth row and jth column block) is given by:

$$\nabla_{\mathbf{w}_k} \nabla_{\mathbf{w}_j} E(\mathbf{w}_1, \ldots, \mathbf{w}_K) = -\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj}) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T$$

To prove the positive semidefinite property, we consider the product $\mathbf{u}^T\mathbf{H}\mathbf{u}$ where $\mathbf{u}$ is an arbitrary vector of length MK.

Let

$$\mathbf{u} = [\mathbf{u}_{0\to(M-1)} \qquad \mathbf{u}_{M\to(2M-1)} \qquad \cdots \qquad \mathbf{u}_{(K-1)M\to(KM-1)}]^T$$

where $\mathbf{u}_{0\to(M-1)}$ consists of the first M elements of the vector $\mathbf{u}$, and so on.

The M elements of the jth section in the product of $\mathbf{u}^T\mathbf{H}$ are given by multiplying the vector $\mathbf{u}$ with all the K blocks of the jth column of $\mathbf{H}$:

$$\left(\sum_{k=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \quad \left(\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj}) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T\right)\right)$$

This section is multiplied by the the M elements of the jth block of $\mathbf{u}$ in the product $\mathbf{u}^T\mathbf{H}\mathbf{u}$, giving us:

$$\left(\sum_{k=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \quad \left(\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj}) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T\right)\right) \mathbf{u}^T_{(jM)\to((j+1)M-1)}$$

The final product will have K such terms, one for each j, giving us:

$$\sum_{j=0}^{K} \left(\sum_{k=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \quad \left(\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj}) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T\right)\right) \left(\mathbf{u}^T_{(jM)\to((j+1)M-1)}\right)$$

which is the full product.

This becomes:

$$= \left(\sum_{k=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \quad \left(\sum_{n=1}^{N} y_{nk}(I_{kj} - y_{nj}) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \right) \right) \left(\sum_{j=0}^{K} \mathbf{u}_{(jM)\to((j+1)M-1)} \right)$$

$$= \left(\sum_{k=0}^{K} \sum_{n=1}^{N} \sum_{j=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \qquad y_{nk}(I_{kj} - y_{nj}) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \qquad \mathbf{u}_{(jM)\to((j+1)M-1)} \right)$$

$$= - \left(\sum_{k=0}^{K} \sum_{n=1}^{N} \sum_{j=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \qquad y_{nk} y_{nj} \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \qquad \mathbf{u}_{(jM)\to((j+1)M-1)} \right)$$
$$+ \left(\sum_{k=0}^{K} \sum_{n=1}^{N} \sum_{j=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \qquad y_{nk} I_{kj} \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \qquad \mathbf{u}_{(jM)\to((j+1)M-1)} \right)$$

Considering the first term:

$$= - \left(\sum_{k=0}^{K} \sum_{n=1}^{N} \sum_{j=0}^{K} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \qquad y_{nk} y_{nj} \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \qquad \mathbf{u}_{(jM)\to((j+1)M-1)} \right)$$

$$= - \left(\sum_{k=0}^{K} \sum_{n=1}^{N} \sum_{j=0}^{K} \left(y_{nk} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \boldsymbol{\phi}_n \right) \left(y_{nj} \boldsymbol{\phi}_n^T \mathbf{u}_{(jM)\to((j+1)M-1)} \right) \right)$$

$$= - \sum_{n=1}^{N} \left(\sum_{k=0}^{K} y_{nk} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \boldsymbol{\phi}_n \right) \left(\sum_{j=0}^{K} y_{nj} \boldsymbol{\phi}_n^T \mathbf{u}_{(jM)\to((j+1)M-1)} \right)$$

Both the terms are equal, giving us:

$$= - \sum_{n=1}^{N} \left(\sum_{k=0}^{K} y_{nk} \mathbf{u}^T_{(kM)\to((k+1)M-1)} \boldsymbol{\phi}_n \right)^2$$

Considering the second term:

$$\left(\sum_{k=0}^{K}\sum_{n=1}^{N}\sum_{j=0}^{K}\mathbf{u}^T_{(kM)\to((k+1)M-1)} \qquad y_{nk}I_{kj}\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T \qquad \mathbf{u}_{(jM)\to((j+1)M-1)}\right)$$

$I_{kj}=1$ when $k=j$, otherwise it is 0, giving us:

$$=\left(\sum_{k=0}^{K}\sum_{n=1}^{N}\mathbf{u}^T_{(kM)\to((k+1)M-1)} \qquad y_{nk}\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T \qquad \mathbf{u}_{(kM)\to((k+1)M-1)}\right)$$

$$=\left(\sum_{k=0}^{K}\sum_{n=1}^{N}y_{nk}\left(\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n\right)\left(\boldsymbol{\phi}_n^T\mathbf{u}_{(kM)\to((k+1)M-1)}\right)\right)$$

$$=\left(\sum_{k=0}^{K}\sum_{n=1}^{N}y_{nk}\left(\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n\right)^2\right)$$

$$=\sum_{n=1}^{N}\sum_{k=0}^{K}y_{nk}\left(\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n\right)^2$$

The sum of the two terms becomes:

$$\sum_{n=1}^{N}\sum_{k=0}^{K}y_{nk}\left(\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n\right)^2-\sum_{n=1}^{N}\left(\sum_{k=0}^{K}y_{nk}\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n\right)^2$$

$$=\sum_{n=1}^{N}\left(\sum_{k=0}^{K}y_{nk}\left(\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n\right)^2-\left(\sum_{k=0}^{K}y_{nk}\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n\right)^2\right)$$

Let $f(x)=x^2$, and $x=\mathbf{u}^T_{(kM)\to((k+1)M-1)}\boldsymbol{\phi}_n$. This gives us:

$$\sum_{n=1}^{N}\left(\sum_{k=0}^{K}y_{nk}f\left(x\right)-f\left(\sum_{k=0}^{K}y_{nk}x\right)\right)$$

$1 \geq y_{nk} \geq 0$, and $\sum_k y_{nk} = 1$ since the sum of predicted probabilities is 1.

Also, $f(x) = x^2$ is a convex function. So we can apply Jensen's inequality to say that:

$$\sum_{k=0}^{K} y_{nk} f(x) \geq f\left(\sum_{k=0}^{K} y_{nk} x\right)$$

which implies that $\mathbf{u}^T \mathbf{H} \mathbf{u} \geq 0$, and therefore, $\mathbf{H}$ is positive semidefinite.