Solution to Exercise 11.16

⋆11.16 Consider the expression

$$-\log\Bigl(\sum_{j=0}^{J}\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)\Bigr), \tag{11.89}$$

which appears in the loglikelihood function (11.35) of the multinomial logit model. Let the vector $\boldsymbol{\beta}^j$ have k_j components, let $k \equiv k_0 + \ldots + k_J$, and let $\boldsymbol{\beta} \equiv [\boldsymbol{\beta}^0 \vdots \ldots \vdots \boldsymbol{\beta}^J]$. The $k \times k$ Hessian matrix $\boldsymbol{H}$ of (11.89) with respect to $\boldsymbol{\beta}$ can be partitioned into blocks of dimension $k_i \times k_j$, $i = 0, \ldots, J$, $j = 0, \ldots, J$, containing the second-order partial derivatives of (11.89) with respect to an element of $\boldsymbol{\beta}^i$ and an element of $\boldsymbol{\beta}^j$. Show that, for $i \neq j$, the (i, j) block can be written as

$$p_i p_j \boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{tj},$$

where $p_i \equiv \exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)/(\sum_{j=0}^{J}\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j))$ is the probability ascribed to choice i by the multinomial logit model. Then show that the diagonal (i, i) block can be written as

$$-p_i(1-p_i)\boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{ti}.$$

Let the k–vector $\boldsymbol{a}$ be partitioned conformably with the above partitioning of the Hessian $\boldsymbol{H}$, so that we can write $\boldsymbol{a} = [\boldsymbol{a}_0 \vdots \ldots \vdots \boldsymbol{a}_J]$, where each of the vectors $\boldsymbol{a}_j$ has k_j components for $j = 0, \ldots, J$. Show that the quadratic form $\boldsymbol{a}^{\top}\boldsymbol{H}\boldsymbol{a}$ is equal to

$$\Bigl(\sum_{j=0}^{J} p_j w_j\Bigr)^2 - \sum_{j=0}^{J} p_j w_j^2, \tag{11.90}$$

where the scalar product w_j is defined as $\boldsymbol{W}_{tj}\boldsymbol{a}_j$.

Show that expression (11.90) is nonpositive, and explain why this result shows that the multinomial logit loglikelihood function (11.35) is globally concave.

The derivative of (11.89) with respect to the vector $\boldsymbol{\beta}^i$ is the column vector

$$\frac{-\exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)}{\sum_{j=0}^{J}\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)}\boldsymbol{W}_{ti}^{\top}. \tag{S11.12}$$

The derivative of the numerator with respect to $\boldsymbol{\beta}^j$, for $j \neq i$, is 0. Therefore, using the definition of p_i, the derivative of (S11.12) with respect to $\boldsymbol{\beta}^j$ is

$$\frac{\exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)\exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^j)}{\bigl(\sum_{i=0}^{J}\exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)\bigr)\bigl(\sum_{j=0}^{J}\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)\bigr)}\boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{tj} = p_i p_j \boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{tj}.$$

Similarly, the derivative of (S11.12) with respect to $\boldsymbol{\beta}^i$ is

$$\begin{aligned}&\frac{-\exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)\sum_{i=0}^{J}\exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)+\exp^2(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)}{\left(\sum_{i=0}^{J}\exp(\boldsymbol{W}_{ti}\boldsymbol{\beta}^i)\right)^2}\boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{ti}\\ &\quad=-p_i(1-p_i)\boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{ti}.\end{aligned}$$

This completes the first part of the exercise.

The quadratic form in which we are interested is

$$\boldsymbol{a}^{\top}\boldsymbol{H}\boldsymbol{a}=\sum_{i=0}^{J}\sum_{j=0}^{J}\boldsymbol{a}_i^{\top}\boldsymbol{H}_{ij}\,\boldsymbol{a}_j,$$

where we obtained formulas for the diagonal and off-diagonal blocks of $\boldsymbol{H}$ in the first part of the exercise. Using these, for $i\neq j$ we have

$$\boldsymbol{a}_i^{\top}\boldsymbol{H}_{ij}\,\boldsymbol{a}_j=p_i\,p_j\,\boldsymbol{a}_i^{\top}\boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{tj}\boldsymbol{a}_j=p_i\,p_j\,w_iw_j,\tag{S11.13}$$

and for $i=j$ we have

$$\boldsymbol{a}_i^{\top}\boldsymbol{H}_{ii}\,\boldsymbol{a}_i=-p_i(1-p_i)\boldsymbol{a}_i^{\top}\boldsymbol{W}_{ti}^{\top}\boldsymbol{W}_{ti}\boldsymbol{a}_i=-p_i(1-p_i)w_i^2.\tag{S11.14}$$

Thus the entire quadratic form is

$$\sum_{j=0}^{J}\sum_{i\neq j}p_i\,p_j\,w_iw_j+\sum_{j=0}^{J}\bigl(p_j^2w_j^2-p_j\,w_j^2\bigr).$$

This is equal to expression (11.90), because

$$\begin{aligned}&\Bigl(\sum_{j=0}^{J}p_jw_j\Bigr)^2-\sum_{j=0}^{J}p_jw_j^2\\ &\qquad=\sum_{j=0}^{J}p_j^2w_j^2+\sum_{j=0}^{J}\sum_{i\neq j}p_i\,p_j\,w_iw_j-\sum_{j=0}^{J}p_jw_j^2\\ &\qquad=\sum_{j=0}^{J}\sum_{i\neq j}p_i\,p_j\,w_iw_j+\sum_{j=0}^{J}\bigl(p_j^2w_j^2-p_jw_j^2\bigr).\end{aligned}$$

This completes the second part of the exercise.

In order to show that expression (11.90) is nonpositive, we can use the Cauchy-Schwartz inequality, expression (2.08). Let the $(J+1)$–vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ have typical elements $p_j^{1/2}$ and $p_j^{1/2}w_j$, respectively. Then

$$\|\boldsymbol{x}\|^2=\sum_{j=0}^{J}p_j=1,\quad \|\boldsymbol{y}\|^2=\sum_{j=0}^{J}p_jw_j^2,\quad\text{and}\ \boldsymbol{x}^{\top}\boldsymbol{y}=\sum_{j=0}^{J}p_jw_j.$$

Thus expression (11.90) can be rewritten as

$$|\boldsymbol{x}^\top\boldsymbol{y}|^2 - \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2.$$

The Cauchy-Schwartz inequality tells us that $|\boldsymbol{x}^\top\boldsymbol{y}|^2 \leq \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2$, and this immediately tells us that expression (11.90) must be nonpositive.

This result shows that the multinomial logit loglikelihood function (11.35) is globally concave, because $\boldsymbol{H}$ is the Hessian of the loglikelihood function as well as the Hessian of expression (11.89). The first term in (11.35) is linear in the parameter vectors $\boldsymbol{\beta}^j$, and so it does not contribute to the Hessian at all.