Let $\boldsymbol{X}$, $\boldsymbol{y}$ represent the feature matrix and target vector. Suppose there are k classes and p features. Let $\boldsymbol{Z}_i$ be the $(k \times p)$-by-k matrix where

$$\boldsymbol{Z}_i = \begin{pmatrix} x_{i1} & 0 & \cdots & 0 \\ x_{i2} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ x_{ip} & 0 & \cdots & 0 \\ 0 & x_{i1} & \cdots & 0 \\ 0 & x_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & x_{ip} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & x_{i1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{ip} \end{pmatrix}$$

Let $\hat{\boldsymbol{\beta}}$ denote logistic regression optimum, put $\boldsymbol{u}_i = \boldsymbol{Z}_i^\top \hat{\boldsymbol{\beta}}$, and let $\boldsymbol{H}$ denote the hessian for logistic regression at $\hat{\boldsymbol{\beta}}$.

1 ALO

We have

$$\begin{aligned} \boldsymbol{H}_{/i} &= \boldsymbol{H} - \boldsymbol{Z}_i \nabla^2 \ell_i \left(\boldsymbol{u}_i\right) \boldsymbol{Z}_i^\top \\ \boldsymbol{g}_{/i} &= \boldsymbol{Z}_i \nabla \ell_i \left(\boldsymbol{u}_i\right) \end{aligned}$$

For ALO, we have

$$\tilde{\boldsymbol{\beta}}_{/i} = \hat{\boldsymbol{\beta}} - \boldsymbol{H}_{/i}^{-1} \boldsymbol{g}_{/i}$$

Put $\boldsymbol{K}_i = \boldsymbol{Z}_i^\top \boldsymbol{H}^{-1} \boldsymbol{Z}_i$. Then

$$
\begin{aligned}
\boldsymbol{Z}_i^\top \tilde{\boldsymbol{\beta}}_{/i} &= \boldsymbol{Z}_i^\top \left[\hat{\boldsymbol{\beta}} - \boldsymbol{H}_{/i}^{-1} \boldsymbol{g}_{/i}\right] \\
&= \boldsymbol{u}_i - \boldsymbol{Z}_i^\top \left[\boldsymbol{H} - \boldsymbol{Z}_i \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{Z}_i^\top\right]^{-1} \left(-\boldsymbol{Z}_i \nabla \ell_i\left(\boldsymbol{u}_i\right)\right) \\
&= \boldsymbol{u}_i + \boldsymbol{Z}_i^\top \left[\boldsymbol{H} - \boldsymbol{Z}_i \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{Z}_i^\top\right]^{-1} \left(\boldsymbol{Z}_i \nabla \ell_i\left(\boldsymbol{u}_i\right)\right) \\
&= \boldsymbol{u}_i + \boldsymbol{Z}_i^\top \left[\boldsymbol{H}^{-1} - \boldsymbol{H}^{-1} \boldsymbol{Z}_i \left(\boldsymbol{K}_i - \nabla^2 \ell_i\left(\boldsymbol{u}_i\right)^{-1}\right)^{-1} \boldsymbol{Z}_i^\top \boldsymbol{H}^{-1}\right] \left(\boldsymbol{Z}_i \nabla \ell_i\left(\boldsymbol{u}_i\right)\right) \\
&= \boldsymbol{u}_i + \boldsymbol{K}_i \left[\boldsymbol{I} - \left(\boldsymbol{K}_i - \nabla^2 \ell_i\left(\boldsymbol{u}_i\right)^{-1}\right)^{-1} \boldsymbol{K}_i\right] \nabla \ell_i\left(\boldsymbol{u}_i\right) \\
&= \boldsymbol{u}_i + \boldsymbol{K}_i \left[\boldsymbol{I} - \left(\nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i - \boldsymbol{I}\right)^{-1} \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i\right] \nabla \ell_i\left(\boldsymbol{u}_i\right) \\
&= \boldsymbol{u}_i + \boldsymbol{K}_i \left[\boldsymbol{I} + \left(\boldsymbol{I} - \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i\right)^{-1} \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i\right] \nabla \ell_i\left(\boldsymbol{u}_i\right) \\
&= \boldsymbol{u}_i + \boldsymbol{K}_i \left(\boldsymbol{I} - \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i\right)^{-1} \left[\left(\boldsymbol{I} - \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i\right) + \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i\right] \nabla \ell_i\left(\boldsymbol{u}_i\right) \\
&= \boldsymbol{u}_i + \boldsymbol{K}_i \left[\boldsymbol{I} - \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \boldsymbol{K}_i\right]^{-1} \nabla \ell_i\left(\boldsymbol{u}_i\right)
\end{aligned}
$$

Put

$$
\begin{aligned}
\boldsymbol{A}_i &= \nabla^2 \ell_i\left(\boldsymbol{u}_i\right) \\
\boldsymbol{G}_i &= \boldsymbol{I} - \boldsymbol{A}_i \boldsymbol{K}_i \\
\tilde{\boldsymbol{u}}_i &= \boldsymbol{u}_i + \boldsymbol{K}_i \boldsymbol{G}_i^{-1} \nabla \ell_i\left(\boldsymbol{u}_i\right)
\end{aligned}
$$

Then we can define ALO as

$$
\mathrm{ALO} = \frac{1}{n} \sum_i \ell_i\left(\tilde{\boldsymbol{u}}_i\right)
$$