Additional file 1 - *Revisiting VERTIGO and VERTIGO-CI: Identifying confidentiality breaches and introducing a statistically sound, efficient alternative*

Appendix A Notation

Table A1 Glossary for general notation conventions

Random variable in $\mathbb{R}$	A	Uppercase Non-italic
Random vector in $\mathbb{R}^p$	$\mathbf{A}$	Uppercase Non-italic bold
Scalar in $\mathbb{R}$	a	Lowercase Italic
Vector in $\mathbb{R}^p$	$\boldsymbol{a}$	Lowercase Italic bold
Vector in $\mathbb{R}^p$ with all components equal to 1	$\mathbf{1}_p$	-
Matrix in $\mathbb{R}^{n\times p}$	$\boldsymbol{A}$	Uppercase Italic bold
Identity matrix in $\mathbb{R}^{n\times n}$	$\boldsymbol{I}_n$	-
Gradient of $f(\boldsymbol{\theta})$ (column vector)	$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$	$\boldsymbol{H}$ for Hessian
$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})\vert_{\boldsymbol{\theta}=\boldsymbol{a}}$	$\nabla_{\boldsymbol{\theta}} f(\boldsymbol{a})$	$\boldsymbol{H}(\boldsymbol{a})$ for Hessian
$\max_{1\leq j\leq p} \vert a_j\vert$	$\vert\vert\boldsymbol{a}\vert\vert_\infty$	Infinite norm
$\sum_{j=1}^p \vert a_j\vert$	$\vert\vert\boldsymbol{a}\vert\vert_1$	ℓ1-norm
$\sqrt{\sum_{j=1}^p a_i^2}$	$\vert\vert\boldsymbol{a}\vert\vert_2$	ℓ2-norm
Diagonal matrix with entries of $\boldsymbol{a}$ on diagonal	$\mathrm{diag}(\boldsymbol{a})$	Dimension $p\times p$ for $\boldsymbol{a}\in\mathbb{R}^p$
Quantity $\cdot$ at iteration t (step count)	$\cdot_{(t)}$	Starts with $\cdot_{(0)}$

Table A2 Glossary for quantities that pertain to the regression settings

Analytical dataset with sample size n	$\mathcal{D}=\{\ldots\}_{i=1}^n$
Covariate vector for *ith* individual	$\boldsymbol{x}_i=[x_{i1},\ldots,x_{ip}]^\top$
Covariate matrix in $\mathbb{R}^{n\times p}$	$\boldsymbol{X}=\begin{bmatrix} x_{11} & \cdots & x_{1p}\\ \vdots & \ddots & \vdots\\ x_{n1} & \cdots & x_{np}\end{bmatrix}=\begin{bmatrix}\boldsymbol{x}_1^\top\\ \vdots\\ \boldsymbol{x}_n^\top\end{bmatrix}$
Gram matrix	$\boldsymbol{\mathcal{K}}=[\boldsymbol{X}\ \mathbf{1}_n]\,[\boldsymbol{X}\ \mathbf{1}_n]^\top$
True (unknown) parameters	$\beta_{0\star},\boldsymbol{\beta}_\star$
Penalized estimate of the parameter (penalty parameter λ)	$\widehat{\beta}_0^\lambda,\widehat{\boldsymbol{\beta}}^\lambda$
Penalized estimate of the parameter (intercept not penalized)	$\widehat{\beta}_{0,\mathrm{alt}}^\lambda,\widehat{\boldsymbol{\beta}}_{\mathrm{alt}}^\lambda$
Log-likelihood	$\ell_n(\beta_0,\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^n\log(\cdot)$
Penalized log-likelihood	$l_n^\lambda(\beta_0,\boldsymbol{\beta})$
Penalized log-likelihood (intercept not penalized)	$l_{n,\mathrm{alt}}^\lambda(\beta_0,\boldsymbol{\beta})$

Table A3 Glossary for quantities specific to the vertical setting

Number of covariate-nodes	K
Number of covariates at covariate-node k	$p^{(k)}$
Covariate matrix at covariate-node k	$\boldsymbol{X}^{(k)}$
Dual parameter estimates	$\widehat{\boldsymbol{\alpha}}^\lambda$
Penalized estimate associated with covariate-node k	$\widehat{\boldsymbol{\beta}}^{\lambda(k)}$

Appendix B VERTIGO and VERTIGO-CI Algorithms

Algorithm 3 provides the original VERTIGO algorithm [1], while algorithm 4 provides the original VERTIGO-CI algorithm [2]. Some adjustments have been made to the representation of the original algorithms using the harmonized notation to facilitate the comparison. We note that constrained optimization technique should be considered when numerically applying the Newton-Raphson steps because the components of the parameter $\boldsymbol{\alpha}$ are constrained in $(0, 1)$.

Algorithm 3 VERTIGO Original [1]

Input Local data in each node $\boldsymbol{X}^{(k)}$, $k = 1, ..., K$, shared response $\boldsymbol{y}$, parameter λ, initial solution $\boldsymbol{\alpha}_{(s)}, s = 0$

Output Global solution $\widehat{\boldsymbol{\alpha}}^{\lambda}$

Procedure:

1. Node k: compute the local matrix $\boldsymbol{\mathcal{K}}^{(k)}$ and send $\boldsymbol{\mathcal{K}}^{(k)}$ to CC.
2. CC: compute the global matrix $\boldsymbol{\mathcal{K}}$.
3. Do:
   (a) Node k: compute $\boldsymbol{e}^{(k)}(\boldsymbol{\alpha}_{(s)})$, and send $\boldsymbol{e}^{(k)}(\boldsymbol{\alpha}_{(s)})$ to CC.
   (b) CC: calculate $\boldsymbol{e}_{(s)} = \sum_{k=1}^{K} \boldsymbol{e}^{(k)}(\boldsymbol{\alpha}_{(s)})$ and $\nabla_{\boldsymbol{\alpha}} J^{\lambda}(\boldsymbol{\alpha}_{(s)})$ (using Equation (10)).
   (c) CC: compute the Hessian matrix $\boldsymbol{H}(\boldsymbol{\alpha}_{(s)})$ (using Equation (11)).
   (d) CC: update $\boldsymbol{\alpha}_{(s)} \leftarrow \boldsymbol{\alpha}_{(s)} - \boldsymbol{H}^{-1}(\boldsymbol{\alpha}_{(s)}) \nabla_{\boldsymbol{\alpha}} J^{\lambda}(\boldsymbol{\alpha}_{(s)})$ and send update $\boldsymbol{\alpha}_{(s)}$ to each node.

   $s \leftarrow s + 1$, until $\boldsymbol{\alpha}_{(s)}$ converges.
4. $\widehat{\boldsymbol{\alpha}}^{\lambda} \leftarrow \boldsymbol{\alpha}_{(s)}$.

Algorithm 4 VERTIGO-CI Original [2]

Input Local data in each node $\boldsymbol{X}^{(k)}$, $k = 1, ..., K$, shared response $\boldsymbol{y}$, parameter λ, initial solution $\boldsymbol{\alpha}_{(s)}, s = 0$

Output Global solution $\widehat{\boldsymbol{\beta}}^{\lambda}$ and associated variances.

Procedure:

1. Node k: compute the local matrix $\boldsymbol{\mathcal{K}}^{(k)}$ and send $\boldsymbol{\mathcal{K}}^{(k)}$ to CC.
2. CC: compute the global matrix $\boldsymbol{\mathcal{K}}$.
3. Do:
   (a) Node k: compute $\boldsymbol{e}^{(k)}(\boldsymbol{\alpha}_{(s)})$, and send $\boldsymbol{e}^{(k)}(\boldsymbol{\alpha}_{(s)})$ to CC.
   (b) CC: calculate $\boldsymbol{e}_{(s)} = \sum_{k=1}^{K} \boldsymbol{e}^{(k)}(\boldsymbol{\alpha}_{(s)})$ and $\nabla_{\boldsymbol{\alpha}} J^{\lambda}(\boldsymbol{\alpha}_{(s)})$ (using Equation (10)).
   (c) CC: compute the modified Hessian matrix $\boldsymbol{H}(\boldsymbol{\alpha}_{(s)}) = \lambda^{-1} \operatorname{diag}(\boldsymbol{y}) \boldsymbol{\mathcal{K}} \operatorname{diag}(\boldsymbol{y}) + C\boldsymbol{I}_n$.
   (d) CC: update $\boldsymbol{\alpha}_{(s)} \leftarrow \boldsymbol{\alpha}_{(s)} - \boldsymbol{H}^{-1}(\boldsymbol{\alpha}_{(s)}) \nabla_{\boldsymbol{\alpha}} J^{\lambda}(\boldsymbol{\alpha}_{(s)})$ and send update $\boldsymbol{\alpha}_{(s)}$ to each node.

   $s \leftarrow s + 1$, until $\boldsymbol{\alpha}_{(s)}$ converges.
4. CC: $\widehat{\boldsymbol{\alpha}}^{\lambda} \leftarrow \boldsymbol{\alpha}_{(s)}$.
5. CC: send $\widehat{\boldsymbol{\alpha}}^{\lambda}$ to each node.
6. Node k, $k \leq K-1$: Calculate $\widehat{\boldsymbol{\beta}}^{\lambda(k)} = \lambda^{-1} \sum_{i=1}^{n} \widehat{\alpha}_i^{\lambda} y_i \boldsymbol{x}_i^{(k)}$. and send $\widehat{\boldsymbol{\beta}}^{\lambda(k)}$ to CC.
7. Node K: Calculate $\widehat{\beta}_0^{\lambda} = \lambda^{-1} \sum_{i=1}^{n} \widehat{\alpha}_i^{\lambda} y_i$ and $\widehat{\boldsymbol{\beta}}^{\lambda(K)} = \lambda^{-1} \sum_{i=1}^{n} \widehat{\alpha}_i^{\lambda} y_i \boldsymbol{x}_i^{(K)}$, and send $\widehat{\beta}_0^{\lambda}$ and $\widehat{\boldsymbol{\beta}}^{\lambda(K)}$ to CC.
8. Client-to-client communication:
   (a) Node k, $k \leq K-1$: compute $\exp\{(\boldsymbol{x}_i^{(k)})^{\top} \widehat{\boldsymbol{\beta}}^{\lambda(k)}\}$ for all $i \in \{1, \ldots, n\}$ and send result to node 1.
   (b) Node K: compute and send $\exp\{\widehat{\beta}_0^{\lambda}\} \exp\{(\boldsymbol{x}_i^{(K)})^{\top} \widehat{\boldsymbol{\beta}}^{\lambda(K)}\}$ for all $i \in \{1, \ldots, n\}$ and send result to node 1.
   (c) Node 1: combine the results using $\exp\{\widehat{\beta}_0^{\lambda} + \boldsymbol{x}_i^{\top} \widehat{\boldsymbol{\beta}}^{\lambda}\} = \exp\{\widehat{\beta}_0^{\lambda}\} \prod_{k=1}^{K} \exp\{(\boldsymbol{x}_i^{(k)})^{\top} \widehat{\boldsymbol{\beta}}^{\lambda(k)}\}$ for all $i \in \{1, \ldots, n\}$, compute $\boldsymbol{V}^{\lambda}$ (using Equation (12)) and send V to all nodes k, $k \in \{2, \ldots, K\}$.
9. Node k, $k \leq K-1$: calculate $(\boldsymbol{X}^{(k)})^{\top} (\boldsymbol{V}^{\lambda})^{1/2}$ and send this quantity to CC.
10. Node K: calculate $[\boldsymbol{X}^{(K)} \, \boldsymbol{1}_n]^{\top} (\boldsymbol{V}^{\lambda})^{1/2}$ and send this quantity to CC.
11. CC: build $[\boldsymbol{X} \, \boldsymbol{1}_n]^{\top} \boldsymbol{V}^{\lambda} [\boldsymbol{X} \, \boldsymbol{1}_n]$ by block (see Algorithm 2 in [2] for details), and invert the result to obtain variance estimates.

Appendix C Equivalence between matrix and expanded notation in dual optimization

To obtain the expression of $J^\lambda(\boldsymbol{\alpha}) = \frac{1}{2\lambda}\boldsymbol{\alpha}^\top \operatorname{diag}(\boldsymbol{y})\boldsymbol{\mathcal{K}}\operatorname{diag}(\boldsymbol{y})\boldsymbol{\alpha} + \sum_{i=1}^n L(\alpha_i)$, it suffices to note that

$$
\begin{aligned}
\boldsymbol{\alpha}^\top \operatorname{diag}(\boldsymbol{y})\boldsymbol{\mathcal{K}}\operatorname{diag}(\boldsymbol{y})\boldsymbol{\alpha} &= \begin{bmatrix}\alpha_1 y_1 & \cdots & \alpha_n y_n\end{bmatrix} \begin{bmatrix}\boldsymbol{x}_1^\top\boldsymbol{x}_1 + 1 & \cdots & \boldsymbol{x}_1^\top\boldsymbol{x}_n + 1 \\ \vdots & \ddots & \vdots \\ \boldsymbol{x}_n^\top\boldsymbol{x}_{1+1} & \cdots & \boldsymbol{x}_n^\top\boldsymbol{x}_n + 1\end{bmatrix} \begin{bmatrix}\alpha_1 y_1 \\ \vdots \\ \alpha_n y_n\end{bmatrix} \\
&= \begin{bmatrix}\sum_{i=1}^n \alpha_i y_i(\boldsymbol{x}_i^\top\boldsymbol{x}_1 + 1) & \cdots & \sum_{i=1}^n \alpha_i y_i(\boldsymbol{x}_i^\top\boldsymbol{x}_n + 1)\end{bmatrix} \begin{bmatrix}\alpha_1 y_1 \\ \vdots \\ \alpha_n y_n\end{bmatrix} \\
&= \sum_{j=1}^n \sum_{i=1}^n \alpha_i y_i(\boldsymbol{x}_i^\top\boldsymbol{x}_j + 1)\alpha_j y_j \\
&= \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j(\boldsymbol{x}_i^\top\boldsymbol{x}_j + 1)\,.
\end{aligned}
$$

Appendix D Mathematical development for the dual problem of the ridge logistic regression maximum log-likelihood problem with no penalty over the intercept

Proposition 1 *Consider any $\mathcal{D} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$, with $y_i \in \{-1, 1\}$, $\boldsymbol{x}_i \in \mathbb{R}^p$ for $i = 1, \ldots, n$ and $n \geq 2$. Assume that $\sum_{i=1}^n \mathbb{I}(y_i = 1) \in \{1, \ldots, n-1\}$ (i.e., there exists at least one i such that $y_i = 1$, and at least one j such that $y_j = -1$). Then, for any $\lambda > 0$,*

there exists a unique solution $(\widehat{\beta}^{\lambda}_{0,alt}, \widehat{\boldsymbol{\beta}}^{\lambda}_{alt})$ *to the optimization problem*

$$\max_{\beta_0 \in \mathbb{R}, \boldsymbol{\beta}=[\beta_1,\ldots,\beta_p]^\top \in \mathbb{R}^p} l^{\lambda}_{n,alt}(\beta_0, \boldsymbol{\beta})\,. \tag{D1}$$

Furthermore, $(\widehat{\beta}^{\lambda}_{0,alt}, \widehat{\boldsymbol{\beta}}^{\lambda}_{alt})$ *satisfies*

$$\widehat{\beta}^{\lambda}_{0,alt} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i \log\{(\widehat{\alpha}^{\lambda}_{i,alt})^{-1} - 1\} - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{alt}\right),$$

$$\widehat{\boldsymbol{\beta}}^{\lambda}_{alt} = \lambda^{-1}\boldsymbol{X}^\top \operatorname{diag}(\boldsymbol{y})\widehat{\boldsymbol{\alpha}}^{\lambda}_{alt}\,,$$

where $\widehat{\boldsymbol{\alpha}}^{\lambda}_{alt} = [\widehat{\alpha}^{\lambda}_{1,alt}, \ldots, \widehat{\alpha}^{\lambda}_{n,alt}]^\top$ *is the unique solution to the following minimization problem:*

$$\min_{\boldsymbol{\alpha}\in(0,1)^n} J^{\lambda}_{alt}(\boldsymbol{\alpha}) \quad s.t. \sum_{i=1}^{n} y_i\alpha_i = 0.$$

Proof of Proposition 1. We begin by showing that the optimization problem in (D1) has a unique finite solution. To do this, we first compute that

$$\nabla_{\boldsymbol{\beta}} l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{y_i \boldsymbol{x}_i}{1 + \exp\{y_i(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta})\}} - \lambda\boldsymbol{\beta}\,,$$

and that

$$\frac{\partial}{\partial\beta_0} l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{y_i}{1 + \exp\{y_i(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta})\}}\,.$$

From this, we obtain that

$$\nabla^2_{\boldsymbol{\beta},\beta_0} l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta})$$
$$= \sum_{i=1}^{n} \frac{-\exp\{y_i(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta})\}}{(1 + \exp\{y_i(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta})\})^2} \begin{bmatrix} \boldsymbol{x}_i\boldsymbol{x}_i^\top & \boldsymbol{x}_i \\ \boldsymbol{x}_i^\top & 1 \end{bmatrix} - \lambda \begin{bmatrix} \boldsymbol{I}_p & 0 \\ 0 & 0 \end{bmatrix}.$$

Hence, $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta})$ is strictly concave in $(\beta_0, \boldsymbol{\beta})$. Therefore, if $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta})$ admits a maximizer, it must be unique. To conclude that the optimization problem in (D1) has a unique solution, it remains to show that $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta})$ indeed has a maximizer.

To establish this, it suffices to show that $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}) \to -\infty$ as $\max\{|\beta_0|, \|\boldsymbol{\beta}\|_\infty\} \to \infty$, where we denote the infinite norm $||\boldsymbol{a}||_\infty := \max_{1\le j\le p} |a_j|$ for $a \in \mathbb{R}^p$. Observe that the log-likelihood term is bounded above by 0. Therefore, if $\|\boldsymbol{\beta}\|_\infty \to \infty$, then regardless of whether $|\beta_0| \to \infty$ or $|\beta_0| < \infty$, the penalty term $\lambda \boldsymbol{\beta}^\top \boldsymbol{\beta}$ dominates, and we have $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}) \to -\infty$.

Now suppose $\|\boldsymbol{\beta}\|_\infty \not\to \infty$ but $\max\{|\beta_0|, \|\boldsymbol{\beta}\|_\infty\} \to \infty$, which implies that $|\beta_0| \to \infty$ while $\|\boldsymbol{\beta}\|_\infty$ remains bounded. Because there is at least one $y_i = 1$ and at least one $y_j = -1$, there exists at least one term in the log-likelihood of the form $\log\left([(1 + \exp\{-y_i(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta})\}]^{-1}\right)$ that diverges to $-\infty$ as $|\beta_0| \to \infty$ (since $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ remains bounded), while the others remain bounded from above by 0. Therefore, $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}) \to -\infty$ in this case as well.

The conclusion of the above discussion is that for any $\lambda > 0$, there exists a unique solution $(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}}, \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})$ to the optimization problem in (D1). This concludes the proof of the first part of the proposition. The rest of the proof is dedicated to show the second part of the proposition.

Since $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta})$ is strictly concave and two times continuously differentiable, $(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}}, \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})$ is a stationary point of $l^{\lambda}_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta})$, i.e.,

$$\nabla_{\boldsymbol{\beta}} l^{\lambda}_{n,\mathrm{alt}}(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}}, \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}) = \sum_{i=1}^{n} \frac{y_i \boldsymbol{x}_i \exp\{-y_i(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})\}}{1 + \exp\{-y_i(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})\}} - \lambda \boldsymbol{\beta} = 0\,,$$

and

$$\frac{\partial}{\partial \beta_0} l^{\lambda}_{n,\mathrm{alt}}(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}}, \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}) = \sum_{i=1}^{n} \frac{y_i \exp\{-y_i(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})\}}{1 + \exp\{-y_i(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})\}}$$
$$= \sum_{i=1}^{n} \left\{ \mathbb{I}(y_i = 1) - \frac{\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}{1 + \exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)} \right\} = 0\,, \tag{D2}$$

where, on the last line, $\mathbb{I}(A)$ denotes the indicator function taking the value 1 if A is true, and 0 otherwise.

Using this result, we will next show that the search space in the optimization problem at (D1) can be narrowed to a compact set (see (D5) below).

From the first equation, we deduce that

$$\|\widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\|_\infty \le \lambda^{-1} \sum_{i=1}^{n} \|\boldsymbol{x}_i\|_\infty\,. \tag{D3}$$

Rearranging the equation for $\frac{\partial}{\partial\beta_0} l^{\lambda}_{n,\mathrm{alt}}(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}}, \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})$ gives

$$\sum_{i:y_i=1} \Big\{1 - \frac{\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}{1+\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}\Big\} = \sum_{i:y_i=-1} \frac{\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}{1+\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}\,. \tag{D4}$$

From the latter equation, we will derive in turns a lower bound and an upper bound for $\widehat{\beta}^{\lambda}_{0,\mathrm{alt}}$. Starting with the upper bound, let $n_1 = \#\{i : y_i = 1\}$ and $n_{-1} = \#\{i : y_i = -1\}$. Since we have assumed $n_1 \geq 1$, we have

$$\begin{aligned}
&\sum_{i:y_i=1} \Big\{1 - \frac{\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}{1+\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}\Big\} \\
&\quad \geq 1 - \max_{i:y_i=1} \frac{\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}{1+\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)} \\
&\quad = 1 - \max_{i:y_i=1} \frac{1}{1+\exp\{-(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}})\}} \\
&\quad \geq 1 - \frac{1}{1+\exp\{-\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} - \|\widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\|_\infty \max_{1\leq i\leq n} \|\boldsymbol{x}_i\|_1\}}\,,
\end{aligned}$$

where $||\boldsymbol{a}||_1 := \sum_{j=1}^p |a_j|$ for $a \in \mathbb{R}^p$. On the other hand, the term on the second line of (D4) satisfies

$$\begin{aligned}
&\sum_{i:y_i=-1} \frac{\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)}{1+\exp\left(\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\right)} \\
&\qquad \leq \frac{n_{-1}}{1+\exp\{-\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} - \|\widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\|_\infty \max_{1\leq i\leq n} \|\boldsymbol{x}_i\|_1\}}\,.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&1 - \frac{1}{1+\exp\{-\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} - \|\widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\|_\infty \max_{1\leq i\leq n} \|\boldsymbol{x}_i\|_1\}} \\
&\qquad \leq \frac{n_{-1}}{1+\exp\{-\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} - \|\widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\|_\infty \max_{1\leq i\leq n} \|\boldsymbol{x}_i\|_1\}}\,,
\end{aligned}$$

which implies that

$$\begin{aligned}
\widehat{\beta}^{\lambda}_{0,\mathrm{alt}} &\geq -\log(n_{-1}) - \|\widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\|_\infty \max_{1\leq i\leq n} \|\boldsymbol{x}_i\|_1 \\
&\geq -\log(n) - p\|\widehat{\boldsymbol{\beta}}^{\lambda}_{\mathrm{alt}}\|_\infty \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty\,,
\end{aligned}$$

where, to obtain the last line, we used the fact that $n_{-1} \leq n$, that for any $\boldsymbol{a} \in \mathbb{R}^p$, $\|\boldsymbol{a}\|_1 \leq p\|\boldsymbol{a}\|_\infty$, and that $\max_{1\leq i\leq n} \|\boldsymbol{x}_i\|_\infty \leq \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty$.

Since from similar arguments we can show that $\widehat{\beta}^\lambda_{0,\text{alt}} \leq \log(n) + p \max_{1\leq i\leq n} \|\boldsymbol{x}_i\|_\infty \|\widehat{\boldsymbol{\beta}}^\lambda_{\text{alt}}\|_\infty$, we conclude using (D3) that

$$|\widehat{\beta}^\lambda_{0,\text{alt}}| \leq \log(n) + \lambda^{-1} p \Big\{ \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty \Big\}^2 .$$

To summarize, we have just shown that the solution $(\widehat{\beta}^\lambda_{0,\text{alt}}, \widehat{\boldsymbol{\beta}}^\lambda_{\text{alt}})$ of (D1) lies in the compact set $\Theta_{\boldsymbol{x},\lambda} = \Theta^0_{\boldsymbol{x},\lambda} \times (\Theta^1_{\boldsymbol{x},\lambda})^p$, with

$$\Theta^0_{\boldsymbol{x},\lambda} = \{\beta \in \mathbb{R} : |\beta| \leq \log(n) + \lambda^{-1} p \Big\{ \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty \Big\}^2 \}$$

and

$$\Theta^1_{\boldsymbol{x},\lambda} = \{\beta \in \mathbb{R} : |\beta| \leq \lambda^{-1} \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty \} .$$

Consequently, the search space $\mathbb{R} \times \mathbb{R}^p$ in the optimization problem at (D1) can be restricted to $\Theta_{\boldsymbol{x},\lambda}$, and the optimization problem at (D1) is equivalent to

$$\max_{\beta_0 \in \Theta^0_{\boldsymbol{x},\lambda}, \boldsymbol{\beta} \in (\Theta^1_{\boldsymbol{x},\lambda})^p} l^\lambda_{n,\text{alt}}(\beta_0, \boldsymbol{\beta}) . \tag{D5}$$

Now from the relationship

$$\begin{aligned} &\min_{\alpha\in(0,1)} \alpha x + (1-\alpha)\log(1-\alpha) + \alpha \log(\alpha) \\ &\qquad = \log\Big(\frac{1}{1+e^{-x}}\Big) , \end{aligned}$$

since for any $\beta_0 \in \Theta^0_{\boldsymbol{x},\lambda}$ and $\boldsymbol{\beta} \in (\Theta^1_{\boldsymbol{x},\lambda})^p$, we have

$$\begin{aligned} |\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}_j| &\leq \log(n) + \lambda^{-1} p \Big\{ \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty \Big\}^2 \\ &\qquad + \lambda^{-1} p \Big\{ \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty \Big\} \max_{1\leq j\leq n} \|\boldsymbol{x}_j\|_\infty \\ &\leq \log(n) + 2\lambda^{-1} p \Big\{ \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty \Big\}^2 \\ &< \log(n) + 2 + 2\lambda^{-1} p \Big\{ \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty \Big\}^2 := c_{n,\boldsymbol{x},\lambda} , \end{aligned}$$

where we used the fact that $|\boldsymbol{\beta}^\top \boldsymbol{x}_j| \le p\|\boldsymbol{\beta}\|_\infty \|\boldsymbol{x}_j\|_\infty \le p\|\boldsymbol{\beta}\|_\infty \sum_{i=1}^n \|\boldsymbol{x}_i\|_\infty$, we conclude that

$$
\begin{aligned}
& l^\lambda_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}) = \min_{\boldsymbol{\alpha} \in (\Theta^2_{\boldsymbol{x},\lambda})^n} l^\lambda_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}, \boldsymbol{\alpha}) \\
& \text{with } l^\lambda_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^n \Big\{ \alpha_i y_i (\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta}) + L(\alpha_i) \Big\} - \frac{\lambda}{2} \boldsymbol{\beta}^\top \boldsymbol{\beta} \\
& \text{and } \Theta^2_{\boldsymbol{x},\lambda} = \Big[\frac{1}{1+\exp(c_{n,\boldsymbol{x},\lambda})}, \frac{1}{1+\exp(-c_{n,\boldsymbol{x},\lambda})} \Big] . \qquad \text{(D6)}
\end{aligned}
$$

We recall $L(\alpha_i) = (1-\alpha_i)\log(1-\alpha_i) + \alpha_i \log(\alpha_i)$. Since the solution $\boldsymbol{\alpha}(\beta_0, \boldsymbol{\beta})$ of the optimization problem $\min_{\boldsymbol{\alpha} \in (\Theta^2_{\boldsymbol{x},\lambda})^n} l^\lambda_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}, \boldsymbol{\alpha})$ satisfies

$$
\alpha_i(\beta_0, \boldsymbol{\beta}) = \frac{\exp(-y_i(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta}))}{1+\exp(-y_i(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta}))} \quad \text{for all } i \in \{1, \dots, n\},
$$

and as the optimal $(\widehat{\beta}^\lambda_{0,\mathrm{alt}}, \widehat{\boldsymbol{\beta}}^\lambda_{\mathrm{alt}})$ are such that

$$
\begin{aligned}
0 &= \sum_{i=1}^n y_i \frac{\exp(-y_i(\widehat{\beta}^\lambda_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^\lambda_{\mathrm{alt}}))}{1+\exp(-y_i(\widehat{\beta}^\lambda_{0,\mathrm{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^\lambda_{\mathrm{alt}}))} \\
&= \sum_{i=1}^n y_i \alpha_i(\widehat{\beta}^\lambda_{0,\mathrm{alt}}, \widehat{\boldsymbol{\beta}}^\lambda_{\mathrm{alt}}) ,
\end{aligned}
$$

(see (D2)), then, the search space $(\Theta^2_{\boldsymbol{x},\lambda})^n$ can be further narrowed to

$$
\Theta^3_{\boldsymbol{x},\lambda} := (\Theta^2_{\boldsymbol{x},\lambda})^n \cap \{\boldsymbol{\alpha} : \sum_{i=1}^n \alpha_i y_i = 0\} . \qquad \text{(D7)}
$$

Since for any $\boldsymbol{\alpha} \in \Theta^3_{\boldsymbol{x},\lambda}$ the function $l^\lambda_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}, \boldsymbol{\alpha})$ satisfies

$$
\begin{aligned}
l^\lambda_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}, \boldsymbol{\alpha}) &:= l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\
&= \sum_{i=1}^n \Big\{ \alpha_i y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} + L(\alpha_i) \Big\} - \frac{\lambda}{2} \boldsymbol{\beta}^\top \boldsymbol{\beta} ,
\end{aligned}
$$

the optimization problem in (D5) can be expressed as

$$
\max_{\beta_0 \in \Theta^0_{\boldsymbol{x},\lambda}, \boldsymbol{\beta} \in (\Theta^1_{\boldsymbol{x},\lambda})^p} l^\lambda_{n,\mathrm{alt}}(\beta_0, \boldsymbol{\beta}) = \max_{\boldsymbol{\beta} \in (\Theta^1_{\boldsymbol{x},\lambda})^p} \Big(\min_{\boldsymbol{\alpha} \in \Theta^3_{\boldsymbol{x},\lambda}} l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \Big) . \qquad \text{(D8)}
$$

Now for any fixed $\boldsymbol{\beta}$ the function $\boldsymbol{\alpha} \mapsto l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is convex, and for any fixed $\boldsymbol{\alpha}$ the function $\boldsymbol{\beta} \mapsto l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is concave. Since $\Theta^1_{\boldsymbol{x},\lambda}$ and $\Theta^3_{\boldsymbol{x},\lambda}$ are compact convex sets, we can apply Sion's

minimax theorem [3] and swap the max and the min in the above equation and conclude that

$$\max_{\boldsymbol{\beta}\in(\Theta^1_{\boldsymbol{x},\lambda})^p}\left(\min_{\boldsymbol{\alpha}\in\Theta^3_{\boldsymbol{x},\lambda}} l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta},\boldsymbol{\alpha})\right) = \min_{\boldsymbol{\alpha}\in\Theta^3_{\boldsymbol{x},\lambda}}\left(\max_{\boldsymbol{\beta}\in(\Theta^1_{\boldsymbol{x},\lambda})^p} l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta},\boldsymbol{\alpha})\right)$$

The inner problem can be solved exactly, by equating to 0 the gradient of $l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta},\boldsymbol{\alpha})$ with respect to $\boldsymbol{\beta}$. The solution $\boldsymbol{\beta}(\boldsymbol{\alpha})$ of this equation is given by the following equation:

$$\boldsymbol{\beta}(\boldsymbol{\alpha}) = \lambda^{-1}\sum_{i=1}^n \alpha_i y_i \boldsymbol{x}_i\,.$$

Plugging the expression of $\boldsymbol{\beta}(\boldsymbol{\alpha})$ into the definition of $l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta},\boldsymbol{\alpha})$ yields $l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta}(\boldsymbol{\alpha}),\boldsymbol{\alpha}) = J^\lambda_{\mathrm{alt}}(\boldsymbol{\alpha})$, and therefore

$$\begin{aligned}\max_{\boldsymbol{\beta}\in(\Theta^1_{\boldsymbol{x},\lambda})^p}\left(\min_{\boldsymbol{\alpha}\in\Theta^3_{\boldsymbol{x},\lambda}} l^\lambda_{n,\mathrm{alt}}(\boldsymbol{\beta},\boldsymbol{\alpha})\right) &= \min_{\boldsymbol{\alpha}\in\Theta^3_{\boldsymbol{x},\lambda}}\left(J^\lambda_{\mathrm{alt}}(\boldsymbol{\alpha})\right)\\ &= \min_{\boldsymbol{\alpha}\in(\Theta^2_{\boldsymbol{x},\lambda})^n}\left(J^\lambda_{\mathrm{alt}}(\boldsymbol{\alpha})\right) \quad \text{s.t.}\sum_{i=1}^n y_i\alpha_i = 0\,.\end{aligned}$$

To obtain the last line, we used the definition of $\Theta^3_{\boldsymbol{x},\lambda}$ in (D7).

The optimization problem on the last line is strongly convex with a linear equality constraint. It therefore has a unique solution provided the feasible set $\{\boldsymbol{\alpha} \in (\Theta^2_{\boldsymbol{x},\lambda})^n : \sum_{i=1}^n \alpha_i y_i = 0\}$ is non empty. This is the case, since, as we have assumed $n_1 \geq 1$ and $n_{-1} \geq 1$ with $n \geq 2$, the point $\boldsymbol{\alpha}^f = (\alpha^f_1,\ldots,\alpha^f_n)^\top$ with $\alpha^f_i = 1/(2n_{-1})$ if $y_i = -1$ and $\alpha^f_i = 1/(2n_1)$ if $y_i = 1$ is feasible. To see why this is the case, recalling the definition of $\Theta^2_{\boldsymbol{x},\lambda}$ in (D6), it suffices to note that the inequality $1/(1+\exp\{x\}) \leq \exp(-x)$ implies that

$$\begin{aligned}\frac{1}{2\max(n_{-1},n_1)} &\geq (2n)^{-1} \geq \exp\{-\log(n)-\log(2)\}\\ &\geq \exp\{-\log(n)-2\} \geq \frac{1}{1+\exp\{\log(n)+2\}}\\ &\geq \frac{1}{1+\exp(c_{n,\boldsymbol{x},\lambda})}\,,\end{aligned}$$

and the inequality $1/2 \leq (1+\exp\{-2\})^{-1}$ implies that

$$\frac{1}{2\min(n_{-1},n_1)} \leq \frac{1}{2} \leq \frac{1}{1+\exp(-2)} \leq \frac{1}{1+\exp(-c_{n,\boldsymbol{x},\lambda})}\,.$$

Combining these relationships shows that $\alpha_i^f \in \Theta^2_{\boldsymbol{x},\lambda}$ for all $1 \leq i \leq n$. The proof of the proposition follows from the fact that, since the solution to the optimization problem

$$\min_{\boldsymbol{\alpha} \in (\Theta^2_{\boldsymbol{x},\lambda})^n} \left(J^\lambda_{\text{alt}}(\boldsymbol{\alpha}) \right) \quad \text{s.t.} \sum_{i=1}^n y_i \alpha_i = 0$$

is reached at a stationary point that cancels the Lagrangian, it can be equivalently expressed as

$$\min_{\boldsymbol{\alpha} \in (0,1)^n} \left(J^\lambda_{\text{alt}}(\boldsymbol{\alpha}) \right) \quad \text{s.t.} \sum_{i=1}^n y_i \alpha_i = 0\,.$$

Finally, the expression $\widehat{\beta}^\lambda_{0,\text{alt}} = \frac{1}{n} \sum_{i=1}^n \left(y_i \log\{(\widehat{\alpha}^\lambda_{i,\text{alt}})^{-1} - 1\} - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^\lambda_{\text{alt}} \right)$ can directly be retrieved using the following:

$$\widehat{\alpha}^\lambda_{i,\text{alt}} = \frac{1}{1 + \exp(y_i(\widehat{\beta}^\lambda_{0,\text{alt}} + \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^\lambda_{\text{alt}}))} \quad \text{for all } i \in \{1, \ldots, n\}.$$

□

Appendix E Alternative reverse-engineering procedure

An alternative procedure can be employed to construct the diagonal matrix $\boldsymbol{V}^\lambda$ at the CC when executing VERTIGO-CI, which enables the CC to follow the steps outlined in the main text and subsequently reverse-engineer the feature data of all individuals. Recall the definition of the diagonal matrix $\boldsymbol{V}^\lambda$, with its entries specified in Equation (12). Also, recall the expression of $\nabla_{\boldsymbol{\alpha}} J^\lambda(\boldsymbol{\alpha})$ (Equation (10)), that $\widehat{\boldsymbol{\alpha}}^\lambda$ satisfies $\nabla_{\boldsymbol{\alpha}} J^\lambda(\widehat{\boldsymbol{\alpha}}^\lambda) = 0$, that $\widehat{\boldsymbol{\beta}}^{\lambda(k)} = \lambda^{-1} \sum_{i=1}^n \widehat{\alpha}^\lambda_i y_i \boldsymbol{x}_i^{(k)}$ and that $\widehat{\beta}^\lambda_0 = \lambda^{-1} \sum_{i=1}^n \widehat{\alpha}^\lambda_i y_i$.

Using these expression, we derive that the jth component of $\nabla_{\boldsymbol{\alpha}} J^\lambda(\boldsymbol{\alpha})$ satisfies

$$
\begin{aligned}
0 &= [\nabla_{\boldsymbol{\alpha}} J^\lambda(\widehat{\boldsymbol{\alpha}}^\lambda)]_j \\
&= \sum_{k=1}^{K} e_j^{(k)}(\widehat{\boldsymbol{\alpha}}^\lambda) + \log\left(\frac{\widehat{\alpha}_j^\lambda}{1-\widehat{\alpha}_j^\lambda}\right) \\
&= \sum_{k=1}^{K} (\lambda^{-1} \sum_{i=1}^{n} \widehat{\alpha}_i^\lambda y_i y_j (\boldsymbol{x}_i^{(k)})^\top \boldsymbol{x}_j^{(k)}) + \lambda^{-1} \sum_{i=1}^{n} \widehat{\alpha}_i^\lambda y_i y_j + \log\left(\frac{\widehat{\alpha}_j^\lambda}{1-\widehat{\alpha}_j^\lambda}\right) \\
&= y_j \sum_{k=1}^{K} ((\boldsymbol{x}_j^{(k)})^\top \widehat{\boldsymbol{\beta}}^{\lambda(k)}) + y_j \widehat{\beta}_0^\lambda + \log\left(\frac{\widehat{\alpha}_j^\lambda}{1-\widehat{\alpha}_j^\lambda}\right) \\
&= y_j (\boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}}^\lambda + \widehat{\beta}_0^\lambda) + \log\left(\frac{\widehat{\alpha}_j^\lambda}{1-\widehat{\alpha}_j^\lambda}\right).
\end{aligned}
$$

By isolating $\widehat{\alpha}_j^\lambda$, we obtain that

$$
\widehat{\alpha}_j^\lambda = \left[1 + \exp\{y_j(\boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}}^\lambda + \widehat{\beta}_0^\lambda)\}\right]^{-1}.
$$

Since $1 - \widehat{\alpha}_j^\lambda = [1 + \exp\{-y_j(\boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}}^\lambda + \widehat{\beta}_0^\lambda)\}]^{-1}$, and as $y_j \in \{-1, 1\}$, we deduce from the latter equation that

$$
\begin{aligned}
&\widehat{\alpha}_j^\lambda (1 - \widehat{\alpha}_j^\lambda) \\
&\quad = \left[1 + \exp\{(\boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}}^\lambda + \widehat{\beta}_0^\lambda)\}\right]^{-1} \left[1 + \exp\{-(\boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}}^\lambda + \widehat{\beta}_0^\lambda)\}\right]^{-1} \\
&\quad = V_{jj}^\lambda .
\end{aligned}
$$

Since $\widehat{\alpha}_j^{\lambda}$ is computed at the CC, the latter can determine each diagonal entry $\boldsymbol{V}_{jj}^{\lambda}$ for all $j \in 1, \ldots, n$ and proceed, as outlined in the main text, to reverse-engineer the individual feature data. Notably, this approach allows the CC to reconstruct the matrix $\boldsymbol{V}^{\lambda}$ without requiring access to the response vector $\boldsymbol{y}$.

References

[1] Li Y, Jiang X, Wang S, Xiong H, Ohno-Machado L. VERTIcal Grid lOgistic regression (VERTIGO). Journal of the American Medical Informatics Association. 2016;23(3):570–579. https://doi.org/10.1093/jamia/ocv146.

[2] Kim J, Li W, Bath T, Jiang X, Ohno-Machado L. VERTIcal Grid lOgistic regression with Confidence Intervals (VERTIGO-CI). AMIA Summits on Translational Science Proceedings. 2021;p. 355–363.

[3] Sion M. On general minimax theorems. Pacific Journal of Mathematics. 1958;8(1):171–176.