Supplementary Note

We briefly summarize some of the data analysis methods mentioned in the text. We denote the two values of the categorical variable (e.g., two stimulus classes in optical and functional imaging, or two response classes in receptive field or classification mapping) by the labels 0 and 1. An instance of the multivariate quantity (an “image”) is denoted by a row vector $x$, consisting of $P$ pixel values. We assume that class $j$ ($j=0$ or 1) is associated with $N_j$ observations of $x$; the $k$th observation is a row vector $x_{k}^{[j]}$. With these conventions, the within-class mean for class $j$ is given by the row vector $\mu_j = \frac{1}{N_j} \sum_{k=1}^{N_j} x_{k}^{[j]}$ and the global mean is $\mu = \frac{N_0 \mu_0 + N_1 \mu_1}{N_0 + N_1}$.

The difference image is the row vector $v_{DL} = \mu_1 - \mu_0$.

Formally, linear regression, the Fisher discriminant, and their extensions seek specific linear functions on images (that is, a rule to be applied to images), not images per se. Operation of a linear function $f$ on an image $x$ can be regarded as the matrix product $xf$ of the row vector $x$ and the column vector $f$. A column vector $f$ can be regarded as a transpose of a row (image) vector, $f = v^T$.

The linear regression method seeks a linear function $f_{LR}$ on the set of images that provides the best prediction of the response (0 or 1). $f_{LR} = v_{LR}^T$, where $v_{LR}$ is the row (image) vector that minimizes $\sum_{j=0}^{1} \sum_{k=1}^{N_j} (x_{k}^{[j]} - v_{LR}^T (x_{k}^{[j]} - \mu))^2$.

$f_{LR} = v_{LR}^T$ may be calculated from the covariance matrix,

$$S = \sum_{j=0}^{1} \sum_{k=1}^{N_j} (x_{k}^{[j]} - \mu)^T (x_{k}^{[j]} - \mu).$$

Provided that the covariance matrix $S$ is invertible, $f_{LR} = \frac{N_0 N_1}{N_0 + N_1} S^{-1} (\mu_1 - \mu_0)^T$. If the covariance matrix is not invertible, $v_{LR}$ is not unique.

The truncated difference method restricts the LR estimate to the subspace spanned by the eigenvectors of $S$ whose eigenvalues are within some range $\lambda_{\text{min}} < \lambda < \lambda_{\text{max}}$. (For a symmetric matrix $M$, a column vector $c$ is said to be an eigenvector of $M$ if $Mc = \alpha c$ for some scalar $\alpha$, and $\alpha$ is said to be the eigenvalue corresponding to $c$).

Ridge regression adds a multiple of the identity to $S$, i.e., replaces $S$ by $S + \kappa I$ in the above. Other regularization procedures replace $S$ by $S + \kappa I + \rho C$, where nonzero elements of $C$ reflect penalties for a lack of smoothness in the estimated image $v$. Here,
κ and ρ are scalars, typically chosen by optimizing the ability of a model based on ν to predict stimulus-response relationships in a separate dataset.

The canonical variates are the solutions of the generalized eigenvalue problem
\[
(\mu - \mu_0)^T (\mu - \mu_0) f = \lambda (S_0 + S_1) f,
\]
where \( S_0 \) and \( S_1 \) are the within-class covariance matrices,
\[
S_j = \sum_{k=1}^{N_j} (x_k^j - \mu_j)^T (x_k^j - \mu_j).
\]
The Fisher discriminant ²⁷ is the linear function \( f_{FD} \) on the set of images that best discriminates between the images that correspond to the two response classes. That is, \( f_{FD} \) maximizes the ratio of the projected difference between classes, \( \left\| (\mu_i - \mu_0) f_{FD} \right\|^2 \), to the projected variances within classes, \( \sum_{j=0}^{N_i} \sum_{k=1}^{N_j} \left\| (x_k^j - \mu_j) f_{FD} \right\|^2 \). For Gaussian data, this maximum is achieved when \( f_{FD} \) is the eigenvector corresponding to the largest eigenvalue of the above generalized eigenvalue problem. Method I of the indicator function method ³⁰ is essentially the Fisher discriminant. Method II considers multiple eigenvectors, whose eigenvalues are sufficiently large. The generalized indicator function method ²⁹ considers eigenvectors of the more general operator
\[
(\mu - \mu_0)^T (\mu - \mu_0) - \alpha (S_0 + S_1)
\]
for some “quality control” parameter \( \alpha \), adds a regularization term, and applies additional criteria to select and weight these eigenvectors.