Iterated LASSO Reveals Highly Distributed and Variable Representations of Faces, Places, and Objects

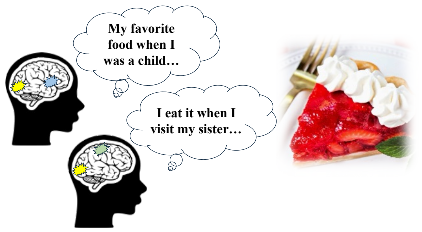
WISCONSIN UNIVERSITY OF WISCONSIN-MADISON

Z.-H. Neuki Li, Y. Ivette Colon, Kushin Mukherjee, Timothy T. Rogers
Department of Psychology, University of Wisconsin-Madison

WISCONSIN INSTITUTE FOR DISCOVERY

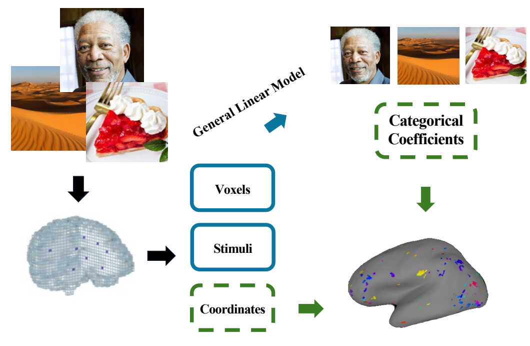
Background

- Recent studies suggested that neural representations may be much more widely distributed and variable than previously suspected (Cox et al., 2023).
- Most decoders are either limited by their implicit assumptions of neural representations or too complex (Frisby et al., 2023).
- We considered iterated LASSO, applying simplistic L1 regularization in an iterative scheme to conduct effective whole-brain voxel feature selection.

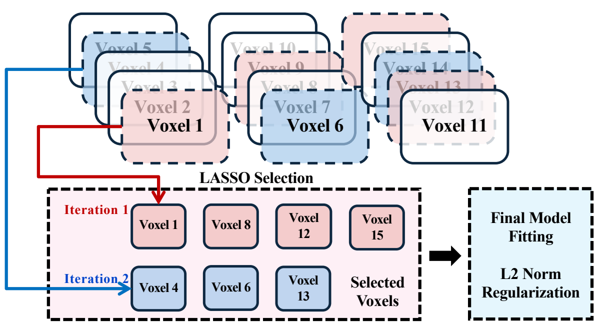


Decoding Neural Representations

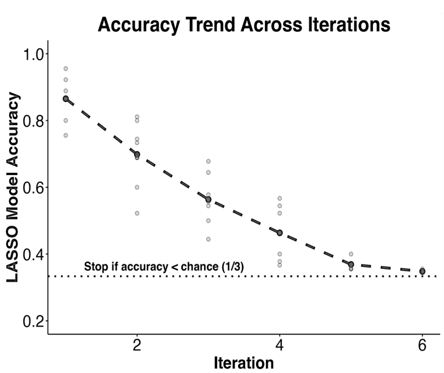
- Functional magnetic resonance imaging (fMRI) data collected when subjects viewing *face*, *place* or *object* visual stimuli. (Lewis-Peacock and Postle, 2008).
- Multi-Variate Pattern Classification (MVPC) conducted by regularized General Linear Models with whole-brain voxel activation data.

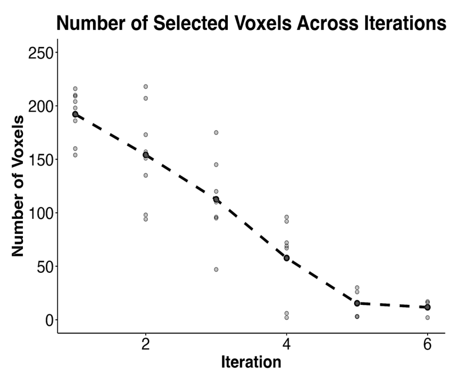


GLM Fitting With Iterated LASSO Feature Selection



- LASSO (L1-norm) regularization penalizes strongly on the coefficients to zero out most of the features in one-round feature selection, mitigating overfitting but also introducing the risk of over-sparsity.
- We applied LASSO iteratively in feature selection to balance the trade-off between overfitting and over-sparsity. Predictivity of voxels are supported by accuracy criterion.





- L2-norm regularization provides milder mitigation on model coefficients.
- We trained final models with all selected voxels with L2-norm regularization to retain all predictive voxels.

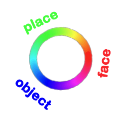
$$\mathcal{L}(\boldsymbol{\beta}) = -\sum_{i=1}^{n}\sum_{k=1}^{3}\mathbb{I}(y_i = k)\cdot\log\left(\frac{\exp\left(\mathbf{x}_i^\top\boldsymbol{\beta}_k\right)}{\sum_{j=1}^{3}\exp(\mathbf{x}_i^\top\boldsymbol{\beta}_j)}\right) \quad \boxed{+\lambda\sum_{k=1}^{3}\|\boldsymbol{\beta}_k\|_1} \quad \boxed{+\lambda\sum_{k=1}^{3}\|\boldsymbol{\beta}_k\|_2^2}$$

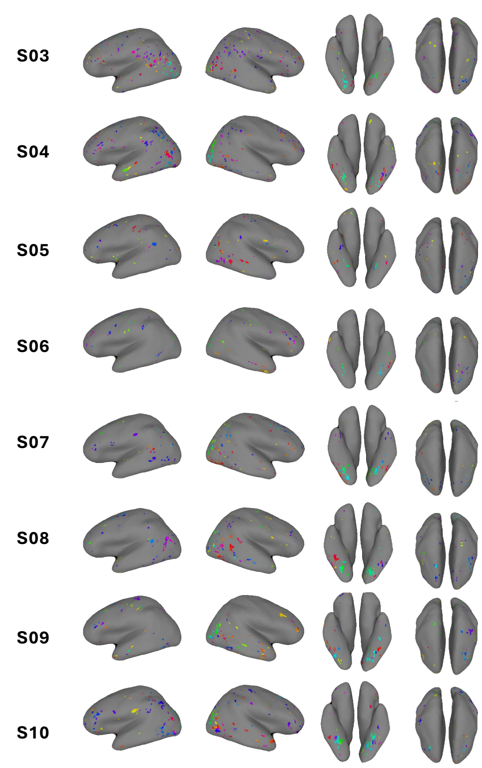
Results

- Individual models decoded stimulus categories from whole-brain fMRI data, achieving 98% mean held-out accuracy.

Subject	Accuracy (%)	# Selected Voxels	Total Voxels	Proportion Selected
3	97.8%	587	4176	14.1%
4	100.0%	648	5651	11.5%
5	97.8%	352	5847	6.0%
6	97.8%	351	8082	4.3%
7	96.7%	435	4101	10.6%
8	95.6%	577	6600	8.7%
9	98.9%	479	9111	5.3%
10	100.0%	645	7193	9.0%

- Predictive voxels found in ~8% area across cortex on average. Many appeared outside the traditional occipito-temporal regions thought to support visual object representation.
- Model weights revealed individual variations in how and where stimulus information is neurally encoded, consistent with prior complex workflows.





Conclusion

- Successfully implemented a simple method with high performance in decoding whole-brain neural representations from fMRI data.
- Balanced between overfitting and over-sparsity with high-dimensional voxel features.
- Revealed distributed and individually variable neural representations of categorical visual stimuli.

Acknowledgement & Key References

Thanks for the collaborations and discussions with the lab members of Knowledge and Concepts Lab on this work.

[1] Cox, C. R et al. (2024). Representational similarity learning reveals a graded multidimensional semantic space in the human anterior temporal cortex. Imaging Neuroscience, 2, 1-22.
[2] Frisby, S. L., Halai, A. D., Cox, C. R., Ralph, M. A. L., & Rogers, T. T. (2023). Decoding semantic representations in mind and brain. Trends in cognitive sciences, 27(3), 258-281.
[3] Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. Journal of Neuroscience, 28(35), 8765-8771.
[4] Friedman J, Tibshirani R, Hastie T (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." Journal of Statistical Software, 33(1), 1–22. doi:10.18637/jss.v033.i01.