Duality of Bures and Shape Distances with Implications for Comparing Neural Representations

Sarah Harvey[1,2] **Brett Larsen**[1,2,3] **Alex Williams**[1,2]

[1] Flatiron Institute, Center for Computational Neuroscience [2] New York University, Center for Neural Science [3] Flatiron Institute, Center for Computational Mathematics

Measuring representational similarity

Comparative analyses are an important tool for understanding complex systems

How do we quantify similarity between ***neural representations***?

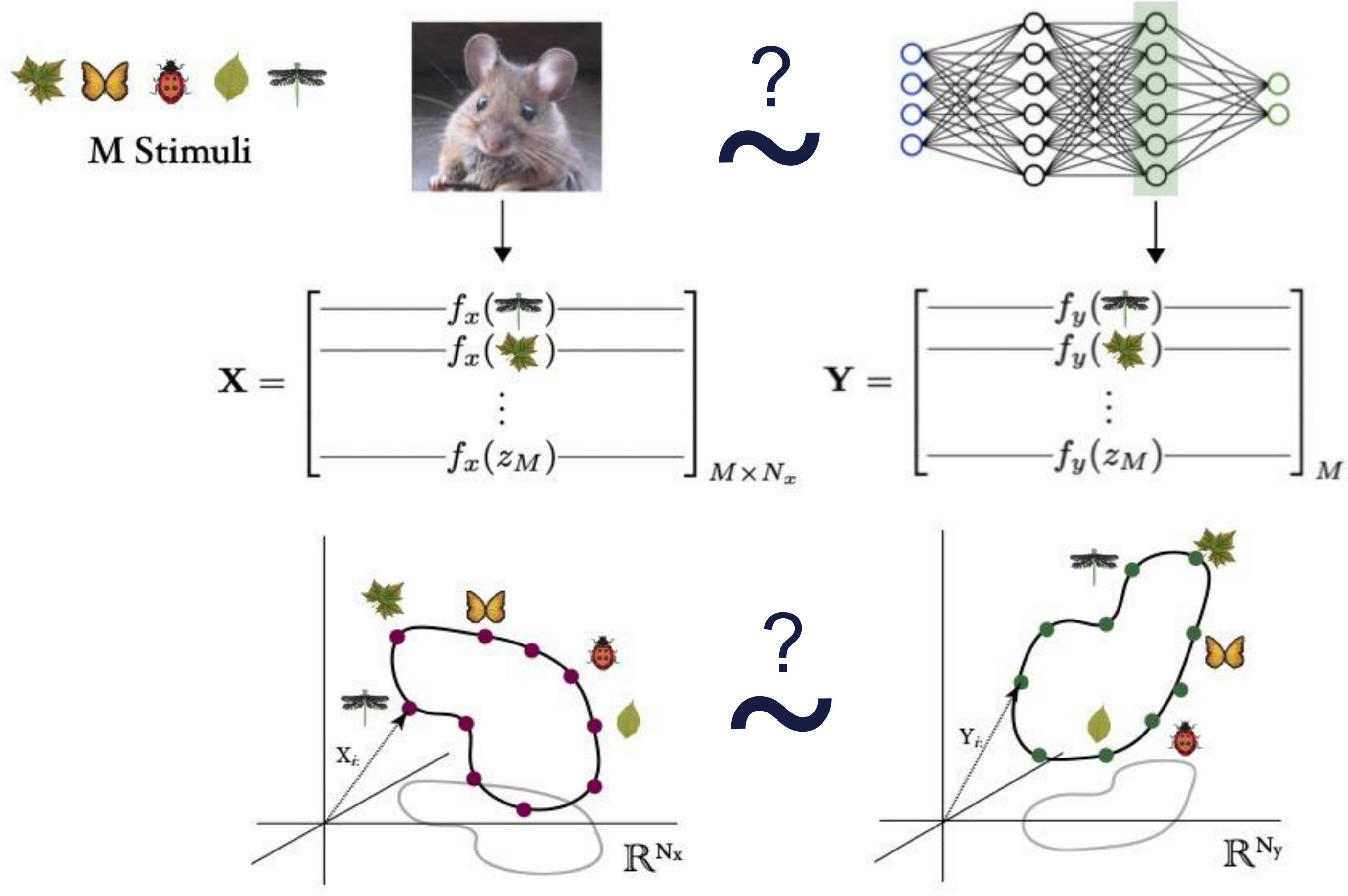
Which method to use depends on what aspects of a representation we believe are important to a system's computations

Ex: For $\mathbf{X} \in \mathbb{R}^{M \times N_x}$ and $\mathbf{Y} \in \mathbb{R}^{M \times N_y}$, we could compute (if $N_x = N_y$):

$$d(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|_F \qquad \textit{Euclidean distance}$$

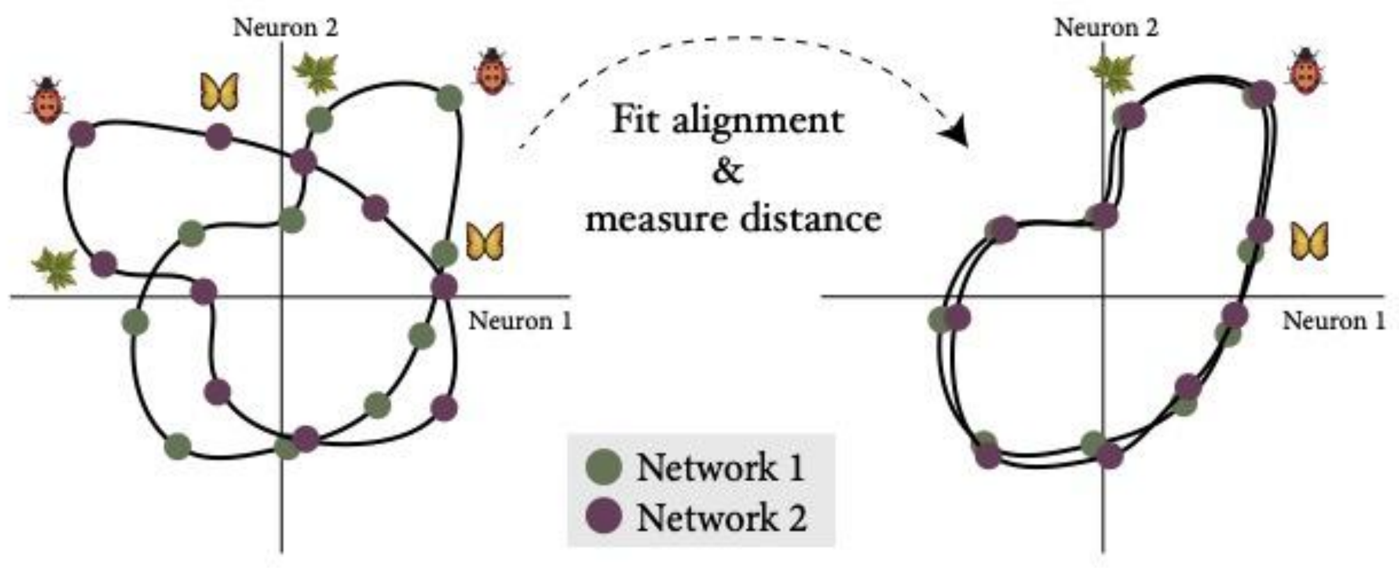
Not invariant to re-indexing of neurons, scalings, etc.

→ Probably not meaningful


M Stimuli
?
~
$\mathbf{X} = \begin{bmatrix} f_x(\cdot) \\ f_x(\cdot) \\ \vdots \\ f_x(z_M) \end{bmatrix}_{M \times N_x}$
$\mathbf{Y} = \begin{bmatrix} f_y(\cdot) \\ f_y(\cdot) \\ \vdots \\ f_y(z_M) \end{bmatrix}_{M \times N_y}$
$X_{i:}$
$\mathbb{R}^{N_x}$
?
~
$Y_{i:}$
$\mathbb{R}^{N_y}$

Two categories of methods

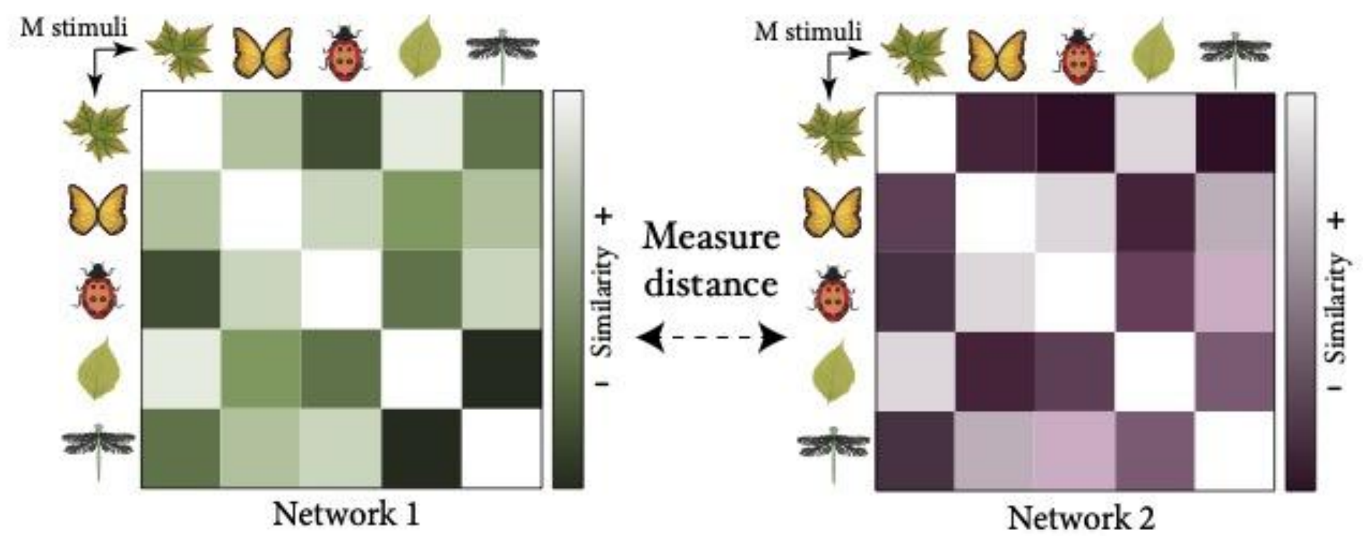
1. (Dis)similarity measures that transform or align neural dimensions



- Fit best nuisance transformation and measure distance between data matrices $\mathbf{X}$ and $\mathbf{Y}$
- Examples:
  - Linear regression [1]
  - Canonical correlations analysis [2]
  - Procrustes shape distances [3]
- Connected to geometry of features in the space of neural activations

$$\underset{g_x, g_y}{\text{minimize}} \;\; d(g_x(\boldsymbol{X}), g_y(\boldsymbol{Y})) \quad \text{subject to} \;\; g_x \in \mathcal{G}_x \;,\; g_y \in \mathcal{G}_y$$

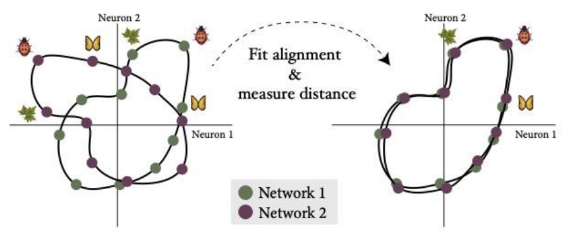
2. (Dis)similarity measures that quantify stimulus-by-stimulus relationships



- Compare M x M (stimulus by stimulus) matrices of summary statistics $\mathbf{K}_X$ and $\mathbf{K}_Y$
- Examples:
  - Representational Similarity Analysis [4].
  - Centered Kernel Alignment (CKA) [5]
  - Bures distance/Normalized Bures Similarity (NBS) [6]
- Connects to cognitive science/psychology literature and history of pairwise similarity experiments

$$d(\mathbf{K}_X, \mathbf{K}_Y); \quad \text{Linear kernel} \;\Rightarrow\; \mathbf{K}_X = \mathbf{X}\mathbf{X}^\top, \; \mathbf{K}_Y = \mathbf{Y}\mathbf{Y}^\top$$

A theoretical link between shape metrics and Bures distance

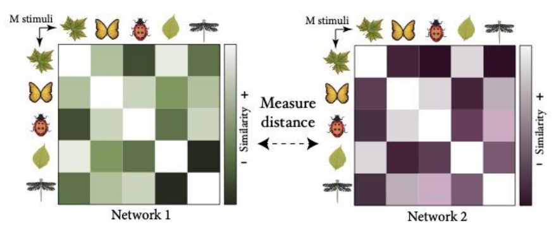


Procrustes size-and-shape distance

$$\mathcal{P}(\mathbf{X}, \mathbf{Y}) = \min_{\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}} \|\mathbf{X} - \mathbf{Y}\mathbf{Q}\|_F$$

Shape similarity

$$\cos\theta^*(\mathbf{X}, \mathbf{Y}) = \max_{\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}} \frac{\mathrm{Tr}[\mathbf{X}^\top\mathbf{Y}\mathbf{Q}]}{\sqrt{\mathrm{Tr}[\mathbf{X}^\top\mathbf{X}]\mathrm{Tr}[\mathbf{Y}^\top\mathbf{Y}]}}$$



Bures distance

$$\mathcal{B}(\boldsymbol{K}_X, \boldsymbol{K}_Y) = \sqrt{\mathrm{Tr}[\boldsymbol{K}_X] + \mathrm{Tr}[\boldsymbol{K}_Y] - 2\,\mathrm{Tr}\left[\left(\boldsymbol{K}_X^{1/2}\boldsymbol{K}_Y\boldsymbol{K}_X^{1/2}\right)^{1/2}\right]}$$

Normalized Bures similarity

$$NBS(\mathbf{K_X}, \mathbf{K}_Y) = \frac{\mathrm{Tr}[(\mathbf{K}_X\mathbf{K}_Y)^{1/2}]}{\sqrt{\|\mathbf{K}_X\|_*\|\mathbf{K}_Y\|_*}}$$

Theorem. $\mathcal{B}(\boldsymbol{K}_X, \boldsymbol{K}_Y) = \mathcal{P}(\boldsymbol{X}, \boldsymbol{Y})$, and furthermore, $NBS(\boldsymbol{K}_X, \boldsymbol{K}_Y) = \cos\theta^*(\boldsymbol{X}, \boldsymbol{Y})$.

⇒ These existing measures of representational similarity are equivalent.

Implications:

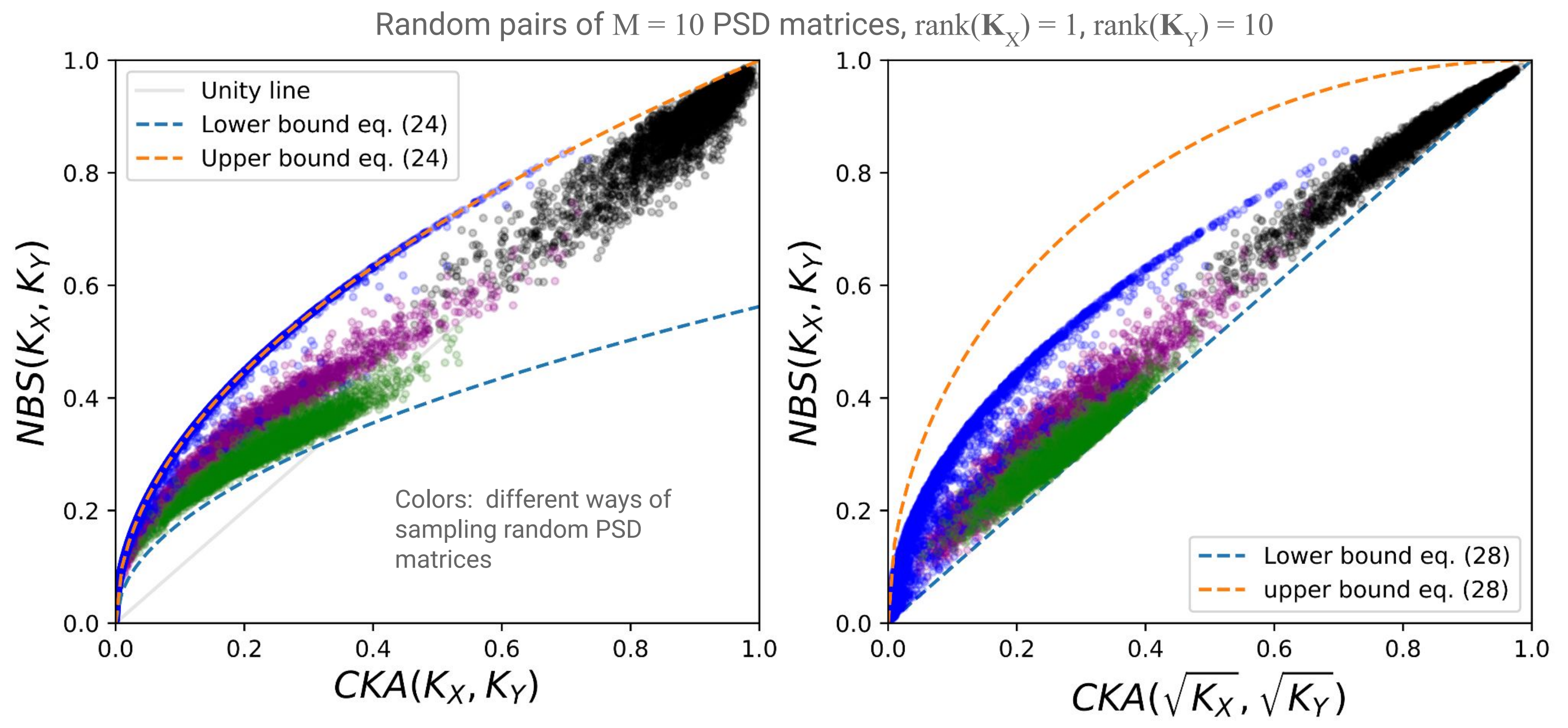
- Provides a way to generalize shape distances to cases where $N_x \neq N_y$
- Provides connections between shape distances and literatures of *optimal transport* and *quantum information theory*
- Enables new insights in continuous cases where $M \to \infty$ or $N_x, N_y \to \infty$.

Comparisons with centered kernel alignment (CKA)

$$CKA(\mathbf{K}_X, \mathbf{K}_Y) = \frac{\mathrm{Tr}[\mathbf{K}_X\mathbf{K}_Y]}{\|\mathbf{K}_X\|_F\|\mathbf{K}_Y\|_F}$$

We find that NBS (or shape similarity) and CKA scores can differ substantially

We derive upper and lower bounds on NBS in terms of CKA by borrowing some results from quantum information theory


Random pairs of M = 10 PSD matrices, $\mathrm{rank}(\mathbf{K}_X) = 1$, $\mathrm{rank}(\mathbf{K}_Y) = 10$

Colors: different ways of sampling random PSD matrices

Bounds in terms of the kernel matrix ranks $r(\cdot)$

$$\frac{\mathrm{CKA}(\boldsymbol{K}_X, \boldsymbol{K}_Y)}{\sqrt{r(\boldsymbol{K}_X)r(\boldsymbol{K}_Y)}} \le \mathrm{NBS}(\boldsymbol{K}_X, \boldsymbol{K}_Y)^2 \le \min[r(\boldsymbol{K}_X), r(\boldsymbol{K}_Y)]\,\mathrm{CKA}(\boldsymbol{K}_X, \boldsymbol{K}_Y)$$

Uhlmann's theorem/Fuchs van de Graaf inequalities

$$1 - NBS(\boldsymbol{K}_X, \boldsymbol{K}_X) \le 1 - CKA(\boldsymbol{K}_X^{1/2}, \boldsymbol{K}_Y^{1/2}) \le \sqrt{1 - NBS(\boldsymbol{K}_X, \boldsymbol{K}_X)^2}$$

Lower rank kernel matrices ⇒ CKA more tightly constrains NBS/shape similarity

Links

Paper:
arXiv:2311.11436
sharvey@flatironinstitute.org

Authors:
@SarahLizHarvey
@_BrettLarsen
@ItsNeuronal

References

1. Yamins, D., DiCarlo, J. Using goal-driven deep learning models to understand sensory cortex. *Nat Neurosci* 19, 356–365 (2016).
2. Raghu, M., et. al. "Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability". NeurIPS 30 (2017).
3. Williams, A. H., et. al.. "Generalized Shape Metrics on Neural Representations". NeurIPS. Vol. 34. (2021).
4. Kriegeskorte N, et. al.. Representational similarity analysis-connecting the branches of systems neuroscience. Front Syst Neurosci. 2008; 2:4.
5. Kornblith, S., Norouzi, M., Lee, H., and Hinton, G. "Similarity of Neural Network Representations Revisited".ICML. Vol. 97. (2019).
6. Muzellec, B., and Cuturi, M. "Generalizing Point Embeddings using the Wasserstein Space of Elliptical Distributions". NeurIPS (2018).