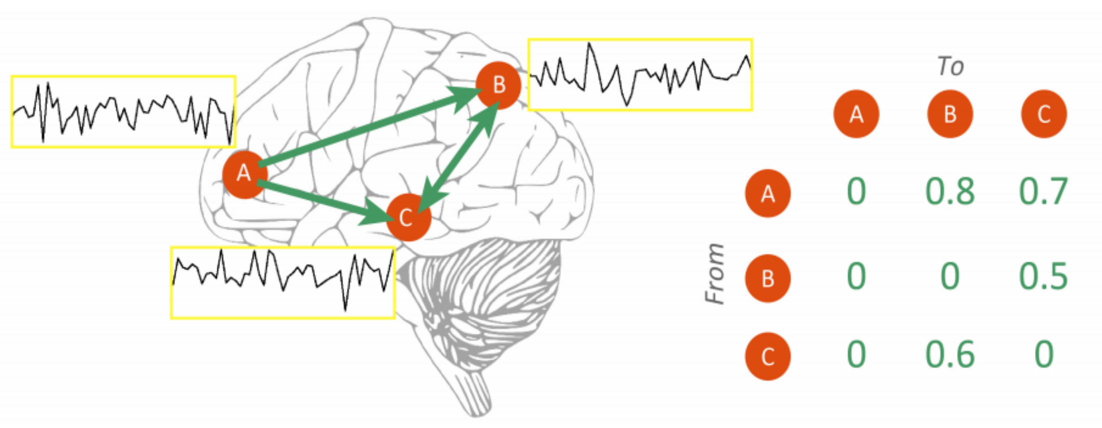
Fingerprinting: measuring information by matching repeated measures

Charles Zheng[1,2], Yuval Benjamini[3], Emily S. Finn[1,4], Natasha Topolski[4,5] and Francisco Pereira[1,2]

1: Functional Magnetic Resonance Imaging Facility, National Institute of Mental Health, 2: Machine Learning Team, National Institute of Mental Health, 3: Department of Statistics and Data Science, Hebrew University of Jerusalem, 4: Section of Functional Imaging Methods, National Institute of Mental Health, 5: McGovern Medical School

Introduction

Functional connectivity (FC) matrices measure the degree of correlation between activation in different brain regions...



and can be used to predict
- diagnosis (e.g. autism)
- age
- fluid intelligence
- memory

... and many more outcomes!

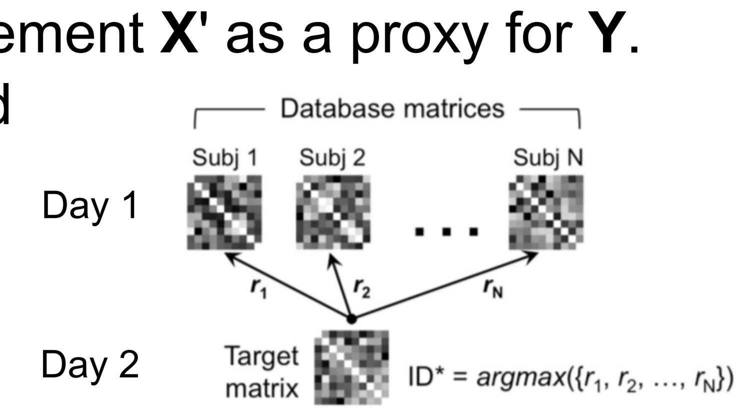
Multiple pipelines produce FC matrices, with many preprocessing choices affecting downstream tasks... creating a methological need for quantitative evaluation of pipelines.

MUTUAL INFORMATION FOR EVALUATING PIPELINE QUALITY

- let **X** denote the FC matrix, and let **Y** denote *all* the **latent subject characteristics** which include or cause the various outcomes of interest (age, diagnosis, etc.)
- mutual information I(**X**; **Y**): an *ideal* criterion that reflects the quality of **X** with regard to the information it contains about **Y**, for *many different* downstream tasks, including prediction.

Our contribution: A novel approach for estimating I(**X**; **Y**) when **Y** is unobserved, by using a repeated measurement **X**' as a proxy for **Y**.

1) Fingerprinting accuracy (FA) [1] defined as accuracy of matching repeated measures from *N* subjects.



2) Applying results from [2] to FA yields an estimator for I(**X**; **X**').

3) **Remaining challenge**: estimating I(**X**; **Y**) from I(**X**; **X**')

We develop a model-based estimation approach (see Theory.)

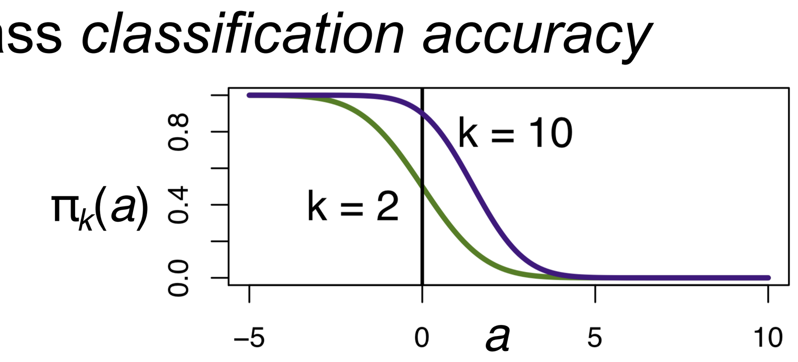
Prior Work: estimating I(X;X')

Estimation of Mutual Information

Classical nonparametric estimators of mutual information do not scale well with dimensionality of **X** and are not practical for neuroimaging data.

Zheng and Benjamini (2016) obtain a high-dimensional estimator of mutual information based on k-class *classification accuracy*

$$\hat{I}(X;Y) = \frac{1}{2}\pi_k^{-1}(1-\text{Acc}))^2.$$



APPLICATION TO FINGERPRINTING

Fingerprinting accuracy, defined

$$\text{Acc} = \frac{1}{n}\sum_{i=1}^{n} I\{i = \text{argmax}_j S_{ij}\}$$

is 1-nearest neighbors classification with feature X and label X'. Therefore, we have

$$\hat{I}(X;X') = \frac{1}{2}\pi_N^{-1}(1-\text{Acc})^2.$$

References

[1]Finn, Emily S., et al. "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity." *Nature neuroscience* 18.11 (2015).

[2]Zheng, Charles Y., and Yuval Benjamini. "Estimating mutual information in high dimensions via classification error." arXiv preprint arXiv:1606.05229 (2016).

[3]Murphy, Kevin, and Michael D. Fox. "Towards a consensus regarding global signal regression for resting state functional connectivity MRI." *Neuroimage* 154 (2017): 169-173.

National Institutes of Health

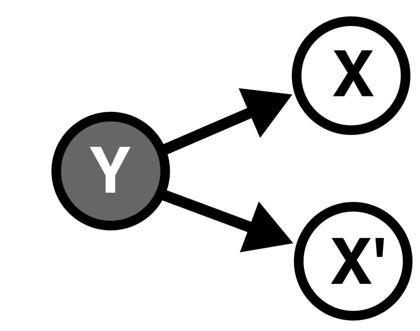
האוניברסיטה העברית בירושלים THE HEBREW UNIVERSITY OF JERUSALEM

FMRIF FIM Machine Learning Team

McGovern MEDICAL SCHOOL

Theory: estimating I(X;Y)

Problem: Given observed **X** and **X**' and unobserved **Y**, where **X** and **X**' are conditionally independent given **Y**, estimate I(**X**; **Y**).



Example: **Y** are latent subject characteristics, **X** and **X**' are FC matrices measured **on the same subject** over 2 different sessions.

Data-Processing Inequality-based estimate: Due to conditional independence of **X**' on **X**, we have I(**X**; **X**') ≤ I(**X**; **Y**). Hence, any estimator of I(**X**; **X**') serves as an estimator of I(**X**; **Y**). However, the Data-Processing Inequality (DPI) bound can be very loose.

Model-based estimator: Assume the following—

(I) *Gaussian copula*: There exists a bijection g() such that for **Z**=g(**X**), we have

$$\text{Cor}(\boldsymbol{Y}, \boldsymbol{Z}) = \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \rho_d \end{pmatrix},$$

where *d* is the number of independent latent components in **Y**.

(II) *Bounded* ρ: we have $\rho_{max} \geq \rho_1 \geq \ldots \geq \rho_d \geq \rho_{min}$.

Example: **Y** are network activation levels; The ρ_d reflect how easy it is to infer each network activation from the measurement **X**. Bounded ρ ↔ components of Y are neither too easy (ρ_{max}) nor difficult (ρ_{min}) to infer from **X**.

DERIVATION OF ESTIMATOR

1) Using assumption I:

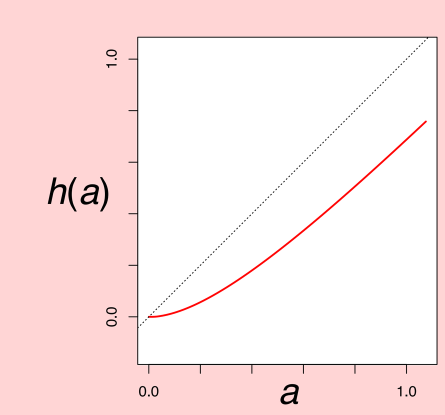
$$I(X;X') = \sum_{i=1}^{d} I(X_i;X_i') = \sum_{i=1}^{d} -\frac{1}{2}\ln(1-\rho_i^4); \qquad I(X;Y) = \sum_{i=1}^{d} I(X_i;Y_i') = \sum_{i=1}^{d} -\frac{1}{2}\ln(1-\rho_i^2).$$

2) Defining $h(a) = -\frac{1}{2}\ln(1-\sqrt{e^{-2a}})$, we have $I(X_i;Y_i) = h(I(X_i;X_i'))$ and hence

$$I(X;Y) = \sum_{i=1}^{d} I(X_i;Y_i) = \sum_{i=1}^{d} h(I(X_i;X_i')).$$

3) Using assumption II, linearly approximate $h(a) \approx \beta_1 a + \beta_0$, hence

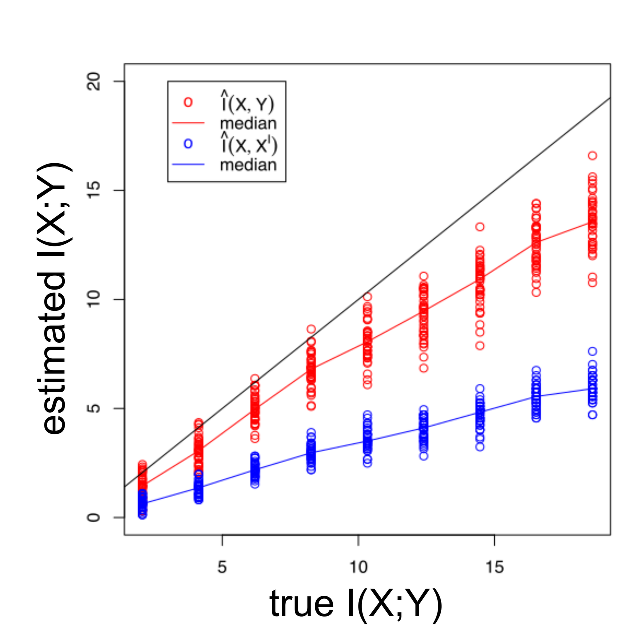
$$I(X;Y) = \sum_{i=1}^{d} h(I(X_i;X_i')) \approx \sum_{i=1}^{d} \beta_1 I(X_i;X_i') + \beta_0 = \beta_1 I(X;X') + d\beta_0.$$



4) *Additional heuristics*: Estimate *d* by $\hat{d} = \dim(X)\frac{\hat{I}(X;X')}{\sum_{i=1}^{p}\hat{I}(X_i;X_i')}$.

Estimate $\rho_{mid} = (\rho_{min} + \rho_{max})/2$ by $\rho_{mid} = \sqrt[4]{1-e^{\frac{-2\hat{I}(X;X')}{\hat{d}}}}$.

Results: DPI vs. model-based



Simulation: (40 repeats per setting)
- **Y** ~ standard multivariate Gaussian
- dim(**Y**) varied from 5 to 45
- $\rho_1 = \ldots = \rho_d = 0.75$
- dim(**X**) = 100, **X** = **AZ** where **A** is a random *dxp* matrix
- N=80 subjects

Results: Diagonal is ideal performance. Both DPI (blue) and model-based (red) underestimate, but model-based is less conservative.

HCP Data: (338 subjects)
- **X** are FC matrices, dim(**X**) ≈ 36,000
- Compared GSR[3] and non-GSR pipelines

Pipeline	Fingerprinting accuracy	DPI estimate I(X; Y)	Estimated d	Model-based estimate I(X; Y)
GSR	0.973	11.4	**75.3**	**27.0**
No-GSR	**0.979**	**12.0**	50.8	24.2

Results: Non-GSR superior with respect to fingerprinting accuracy and DPI estimate. GSR scores higher for model-based I(X;Y) due to higher latent dimensionality.