



The Waisman Laboratory
for Brain Imaging and Behavior



University of Wisconsin
**SCHOOL OF MEDICINE
AND PUBLIC HEALTH**

BMI/STAT-768 Statistical Methods for Medical Image Analysis

Moo K. Chung
Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin-Madison

www.stat.wisc.edu/~mchung

Similarity Measures in Hilbert space

Unit Objectives

- 1) Understand correlations and other correlation-based similarity measures.
- 2) Understand how to compute similarity measures for various data types.

Similarity measures

- Correlation based
- Statistical (goodness-of-fit)
- Information theoretic (mutual information/entropy)
- Geometric (kappa index)

Correlations

- Pearson correlation
- Rank correlation
- Partial correlation
- Multiple correlation
- Partial multiple correlation
- Canonical correlation

Pearson product moment correlation coefficient

Given paired measurements $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, the Pearson product-moment correlation coefficient (Fisher, 1915) is a measure of association given by

$$r_P = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

where \bar{X} and \bar{Y} are the sample mean of X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_n respectively.

Why correlation is **inner product** (vector multiplication)?

Scale and translation invariant

$$r_P(X, Y) = r_P(aX + b, cY + d)$$

Can be written as a matrix multiplication

$$r_P(X, Y) = Z^\top W$$

Question. Prove the above statements

Inference on correlation

$$H_0 : \rho = 0 \text{ vs. } H_1 : \rho \neq 0$$

Under normality,

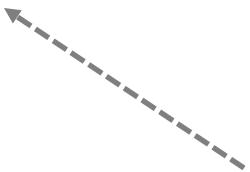
$$T = \frac{r_P \sqrt{n - 2}}{\sqrt{1 - r_P^2}}$$

T-statistic with $n-2$ degrees of freedom

If we can not assume normality of data

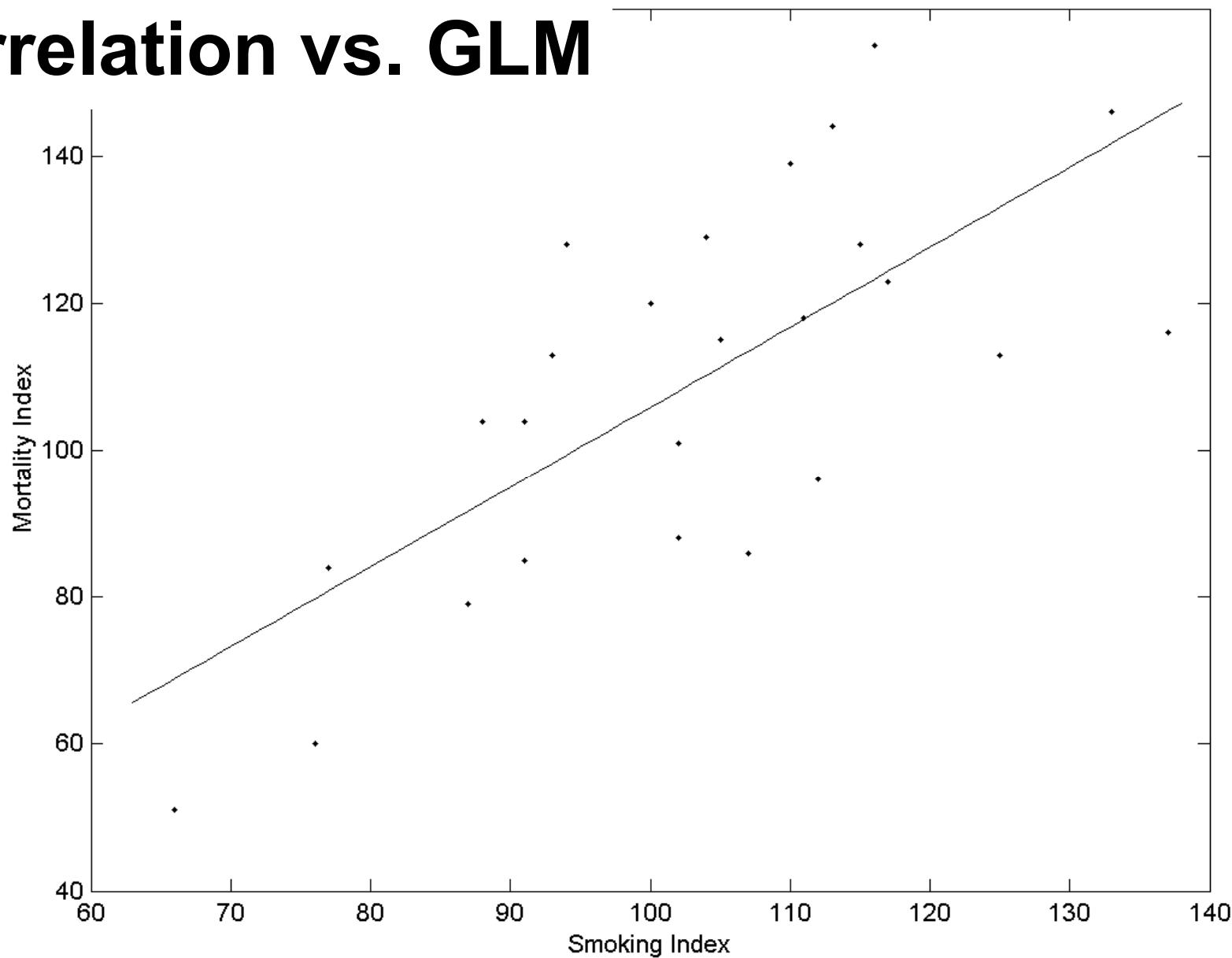
Fisher transform:
$$F(r_P) = \frac{1}{2} \ln \left(\frac{1 + r_P}{1 - r_P} \right)$$

$$Z = \sqrt{n-3} (F(r_P) - F(\rho)) \text{ : standard normal}$$



Note: under the null hypothesis, this term vanishes.

Correlation vs. GLM



Are there any relationship between correlation analysis and GLM ?

Slope of regression vs. correlation

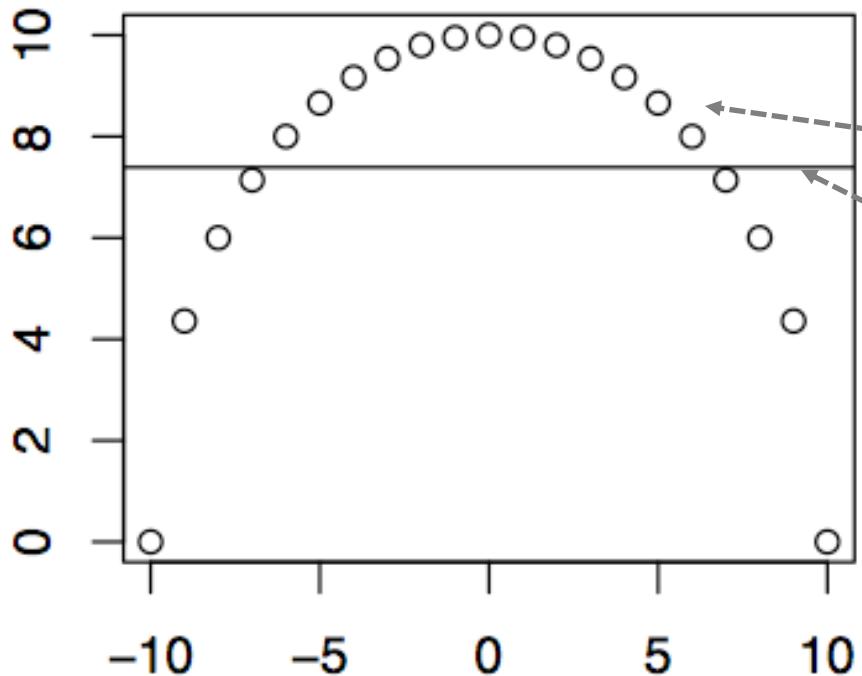
$$Y = \beta_0 + \beta_1 \cdot X$$

If (X, Y) are bivariate normal,

$$\beta_1 \cong \frac{\sigma_Y}{\sigma_X} r_P$$

Under normality, correlation analysis = linear regression

Limitation of correlations



Data $x_i^2 + y_i^2 = 10^2$

Linear regression fit

Slope=correlation=0

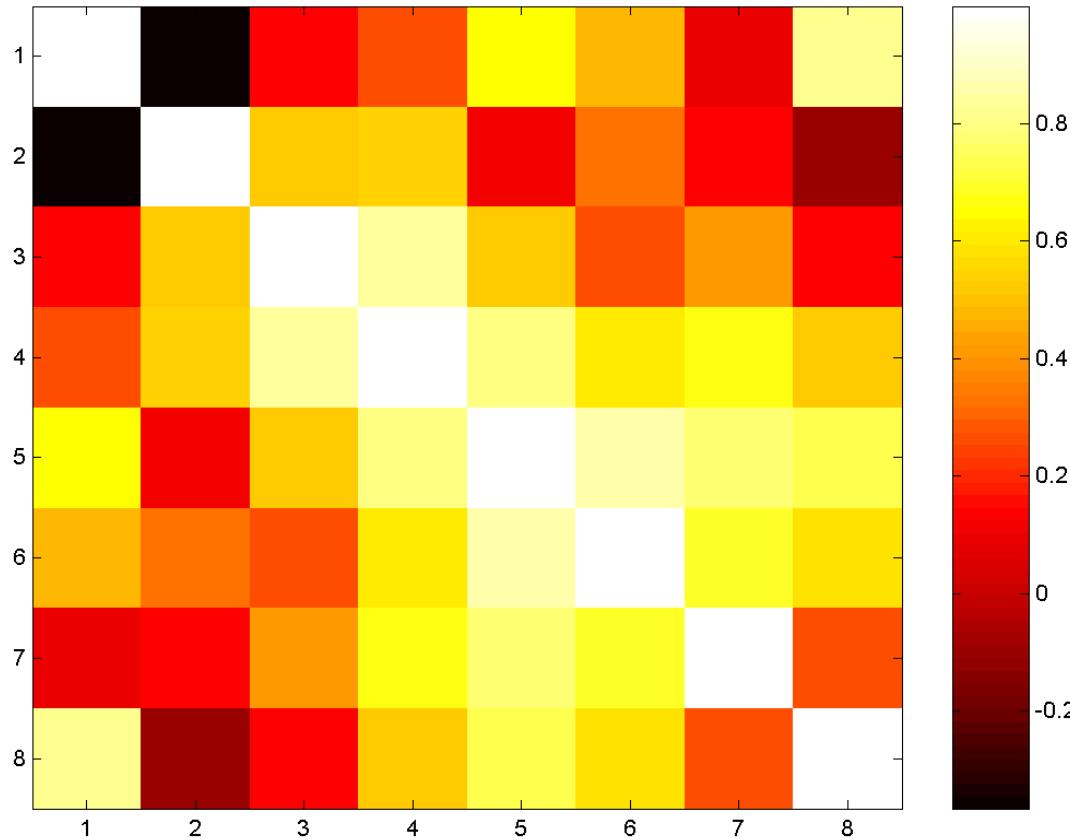
Correlation only measures linear dependency.
It does *not* measure nonlinear dependency.

Question: Mathematically show the correlation = 0

Using correlation, is it possible to measure nonlinear dependency?

Correlation mapping technique

CT intensity values for 8 regions of interests (ROI) in mouse breast cancer study



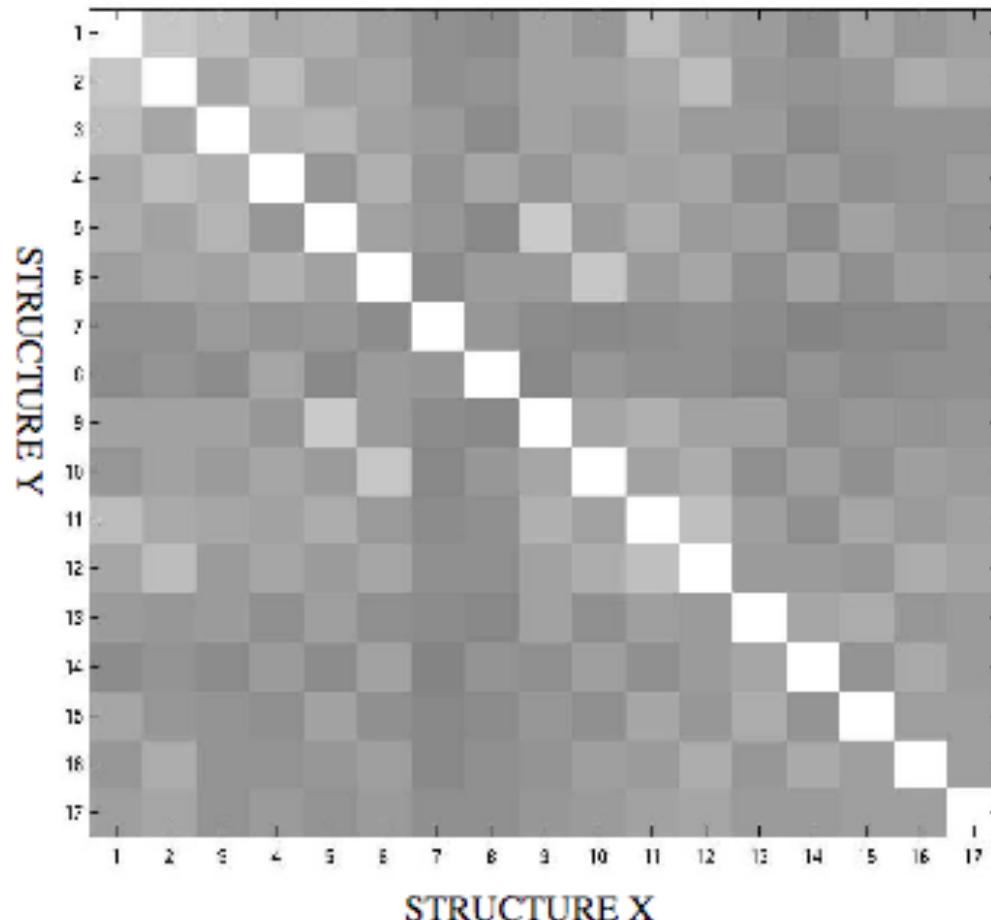
Regions denoted by numbers

Sorted correlation map
can be used to display the
pairwise correlation

→ Dendrogram ([TDA](#))

Correlation mapping technique (anatomical connectivity analysis)

1 = Left Lateral Ventricle
2 = Right Lateral Ventricle
3 = Left Caudate
4 = Right Caudate
5 = Left Putamen
6 = Right Putamen
7 = Left Accumbens
8 = Right Accumbens
9 = Left Pallidum
10 = Right Pallidum
11 = Left Thalamus
12 = Right Thalamus
13 = Left Amygdala
14 = Right Amygdala
15 = Left Hippocampus
16 = Right Hippocampus
17 = Brain Stem



Example

Interested in knowing the relation between

Y_1 = brain cortical thickness

Y_2 = behavioral measure

Biological facts:

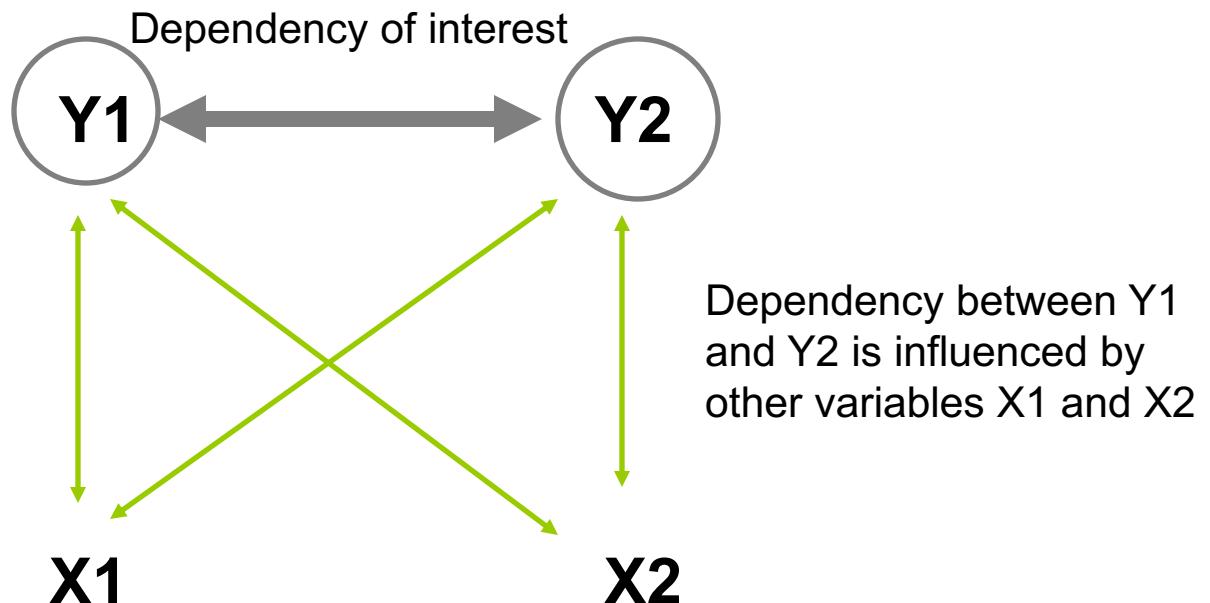
- Cortical thickness is related to age (X_1)
- Cortical thickness is related to brain size (X_2)

We need:

$\text{Corr}(Y_1, Y_2)$ while accounting for *confounding nuisance covariates* X_1 and X_2

Partial correlation

Measure of dependency while removing the effect of other variables.



$$Y = (Y_1, Y_2), X = (X_1, X_2)$$

Covariance matrix: $\mathbb{V}(Y, X)' = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}$

Partial covariance of Y given X:

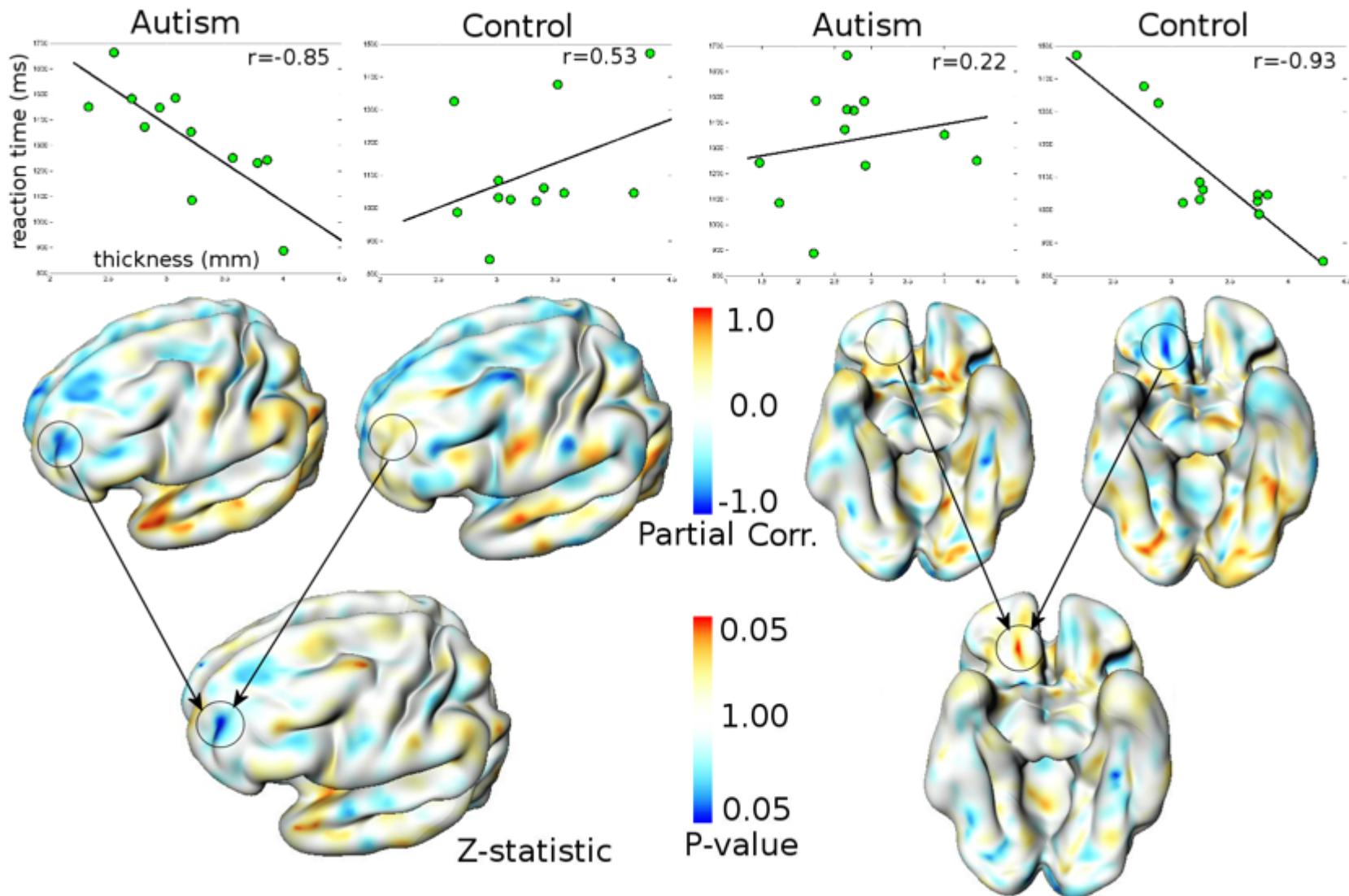
$$\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} = (\sigma_{ij})$$

Under normality,
this is equivalent to the conditional covariance:

$$\text{Cov}(Y|X)$$

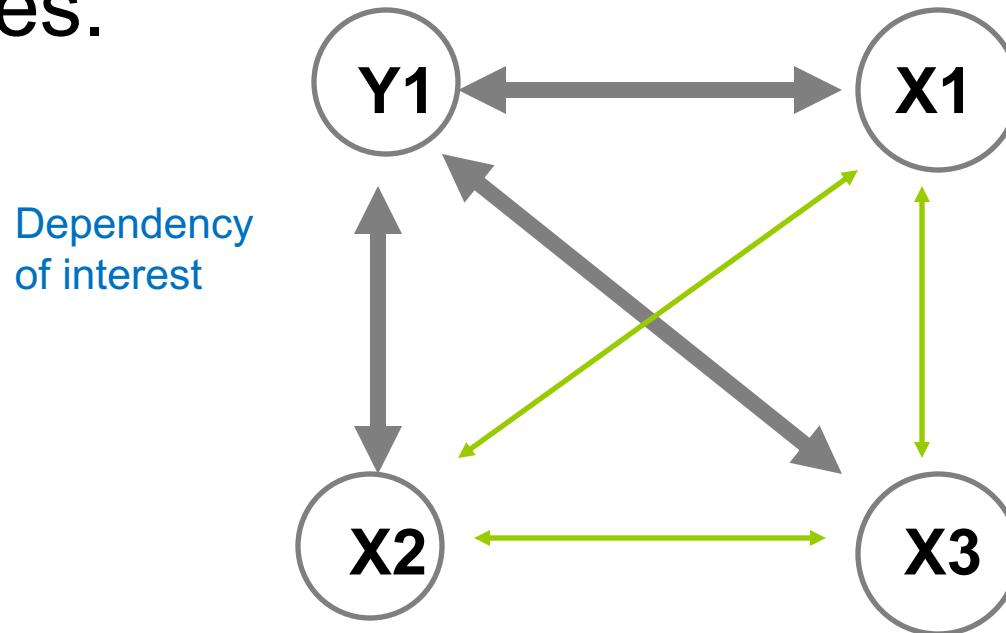
Exercise: prove the above claim

Correlation of emotion recognition response time and cortical thickness while accounting for age and brain size



Multiple correlation

Measure of dependency between one variable and remaining multiple variables.



Example. correlating a reference fMRI time series (amygdala) to fMRI time series in other regions simultaneously

Definition of multiple correlation

Given X_1, X_2, X_3 , consider weighted average $w_1X_1 + w_2X_2 + w_3X_3$

The multiple correlation is defined as

$$\rho = \max_{w_1, w_2, w_3} \text{corr}(Y, w_1X_1 + w_2X_2 + w_3X_3)$$

The maximum of all possible linear combinations of correlations

How to compute multiple correlations?

Least squares estimation

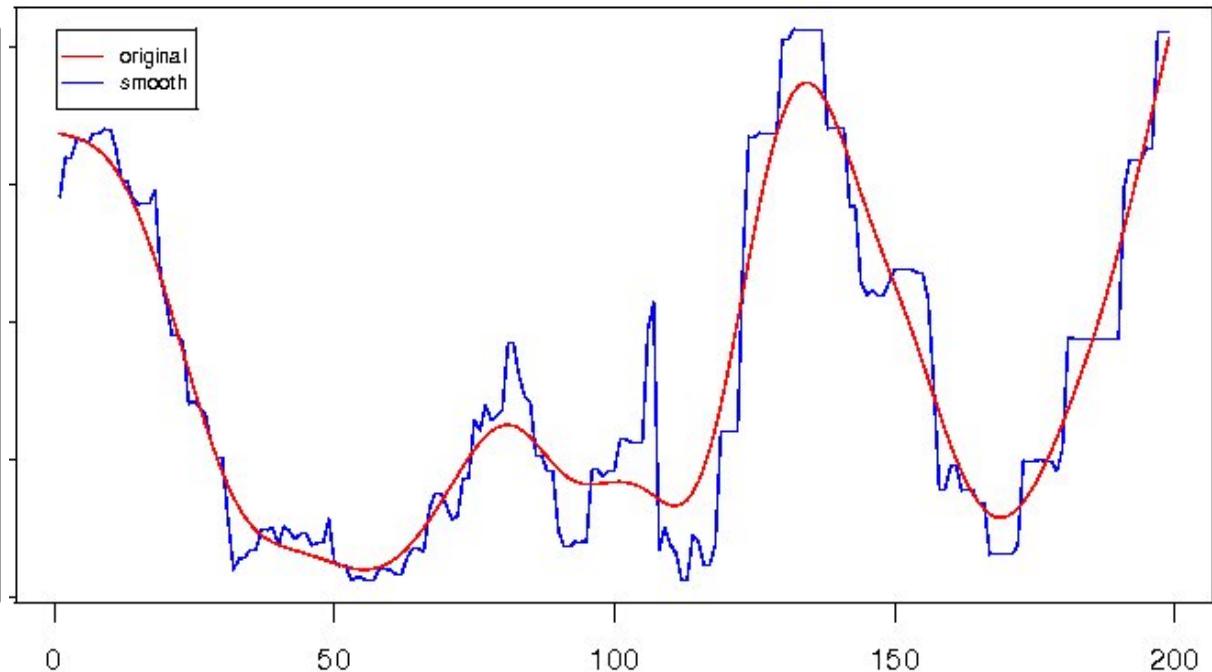
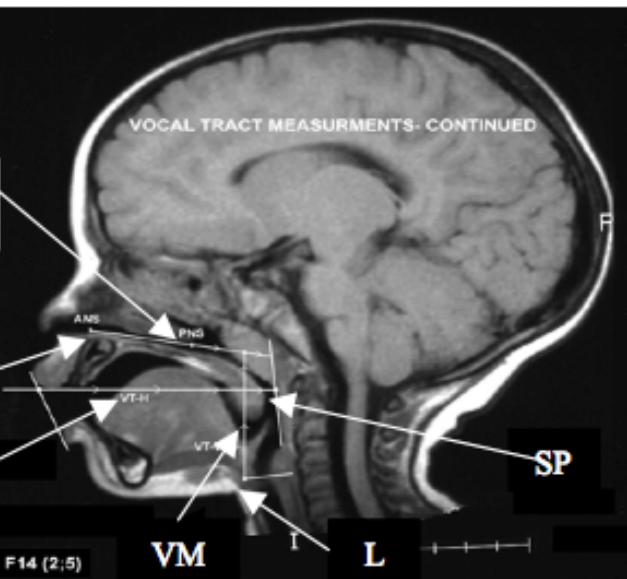
Multiple correlation can be obtained
from multiple linear regression analysis

$$Y = w_1X_1 + w_2X_2 + w_3X_3 + \varepsilon$$

Exercise: Show that LSE gives the maximum correlation

Coordinated growth of vocal tract and other anatomy

Multiple correlation curve of HardP + AntTong + MAndL



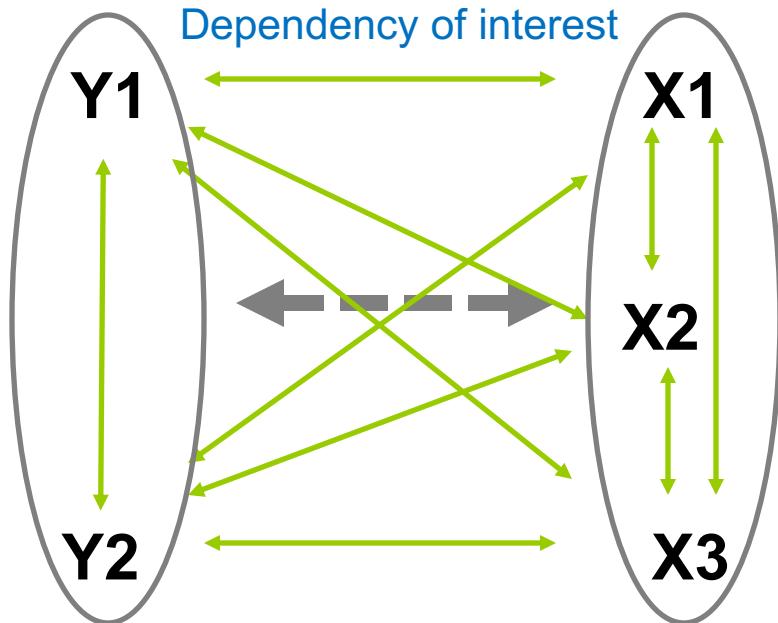
Question.

How anatomical parts develop together longitudinally?
How variables dynamically change together?

Shubing Wang

Canonical correlation

Measure of similarity/dependency between two vectors of unequal size



Applications. network modeling, anatomical and functional connectivity in brain. Distance between two graphs of unequal size

Canonical correlation

$$X = w_1 X_1 + w_2 X_2 + w_3 X_3$$

$$Y = v_1 Y_1 + v_2 Y_2$$

$$\rho = \max_{w,v} \text{corr}(X,Y)$$

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} w = \rho^2 w$$

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} v = \rho^2 v$$

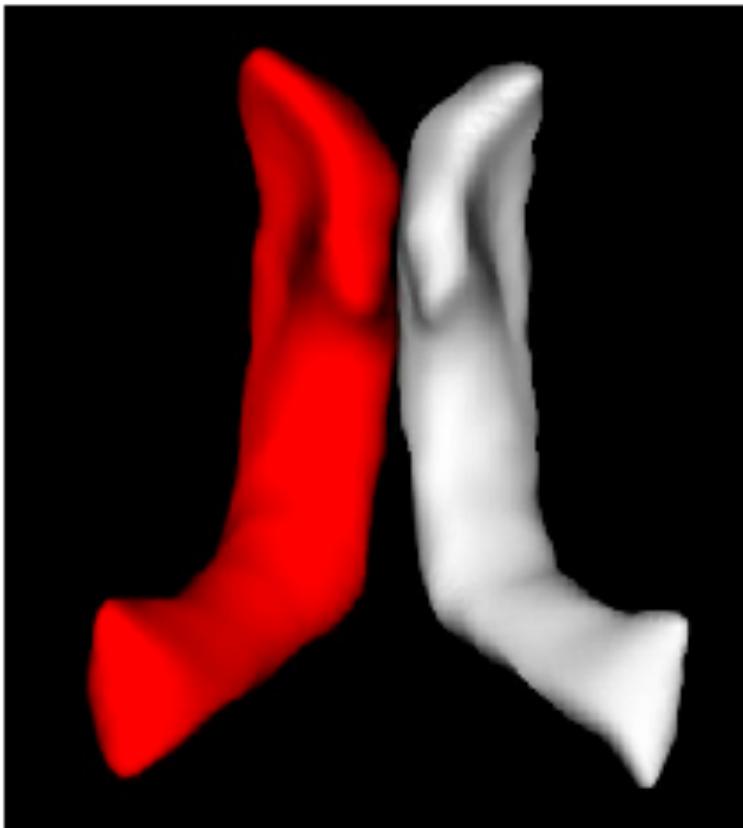


Eigenvalues are identical. The principal eigenvalue is the canonical correlation

Need to solve the eigenequations simultaneously.

Project. Can you develop sparse canonical correlations?

Application to anatomical shape analysis



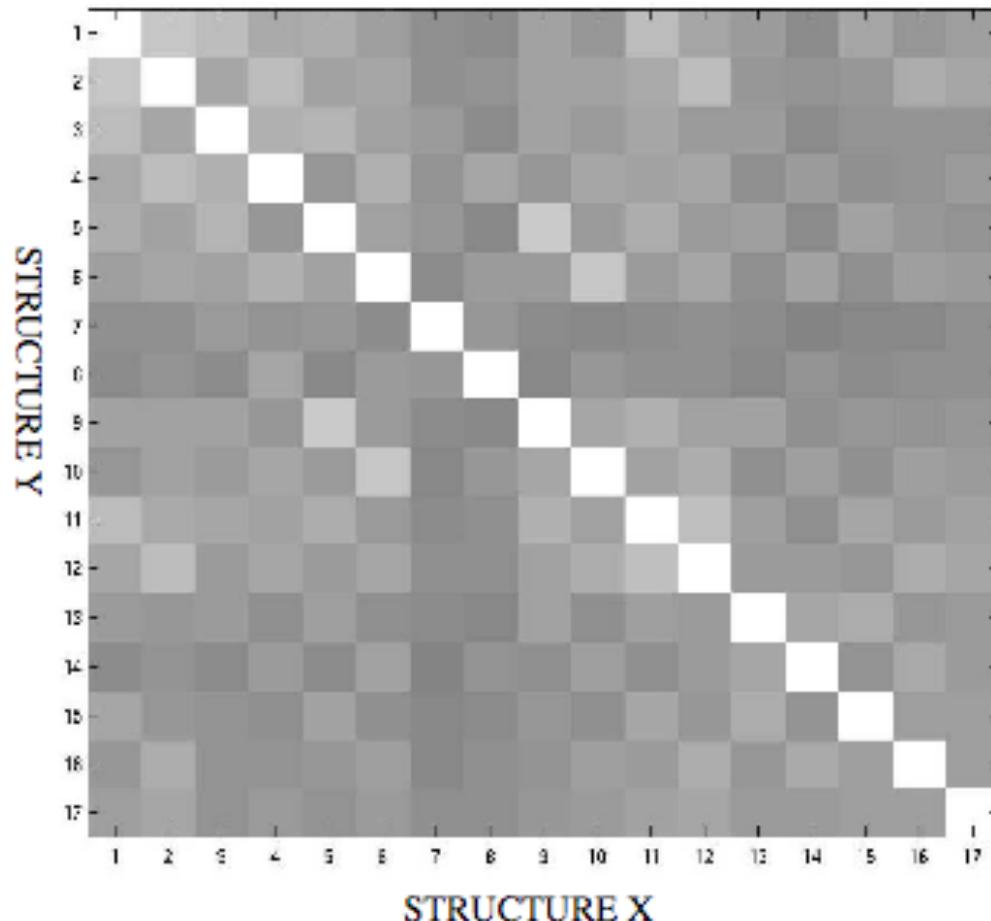
Ventricle surface in brain

Surface coordinates are taken as a vector data
 $x=(x_1, x_2, \dots, x_n)$.

The vector data is canonically correlated with the surface coordinates of other anatomical objects.

Canonical correlation map showing pairwise correlation

- 1 = Left Lateral Ventricle
- 2 = Right Lateral Ventricle
- 3 = Left Caudate
- 4 = Right Caudate
- 5 = Left Putamen
- 6 = Right Putamen
- 7 = Left Accumbens
- 8 = Right Accumbens
- 9 = Left Pallidum
- 10 = Right Pallidum
- 11 = Left Thalamus
- 12 = Right Thalamus
- 13 = Left Amygdala
- 14 = Right Amygdala
- 15 = Left Hippocampus
- 16 = Right Hippocampus
- 17 = Brain Stem

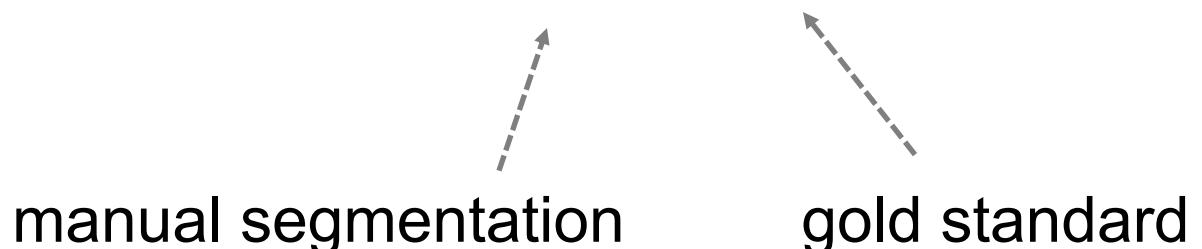


Set theoretic similarity measures

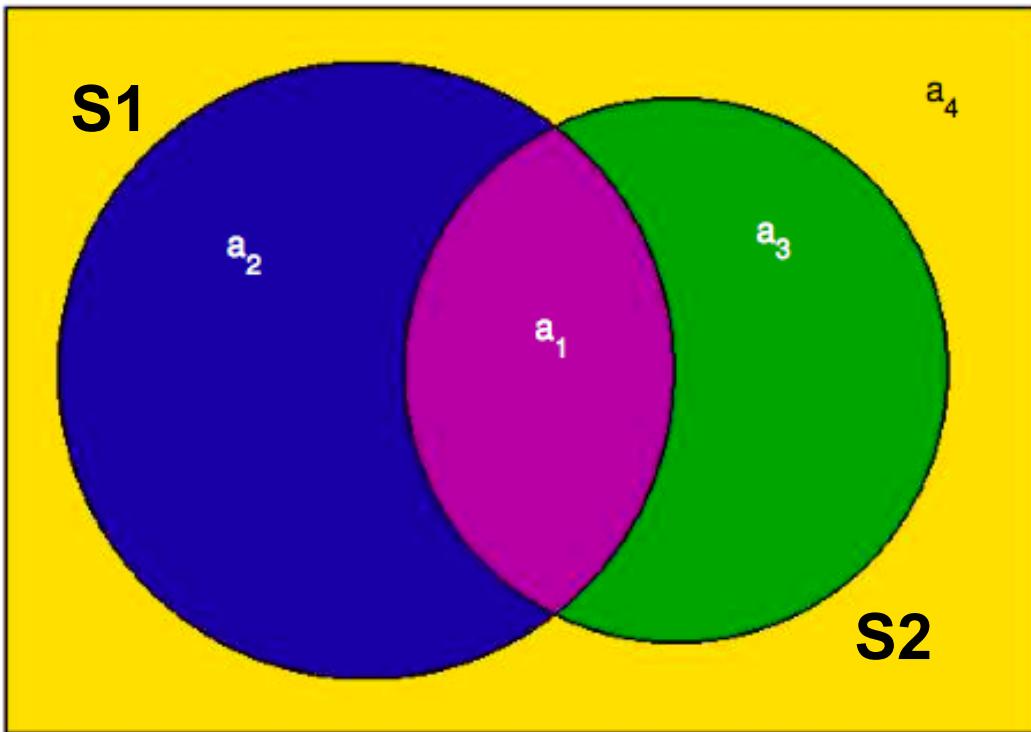
Often used validation metric on mostly on image segmentation and clustering

Kappa index measures the proportion of areal overlap:

$$\kappa(S_1, S_2) = 2 \frac{|S_1 \cap S_2|}{|S_1| + |S_2|}$$



Range of Kappa index

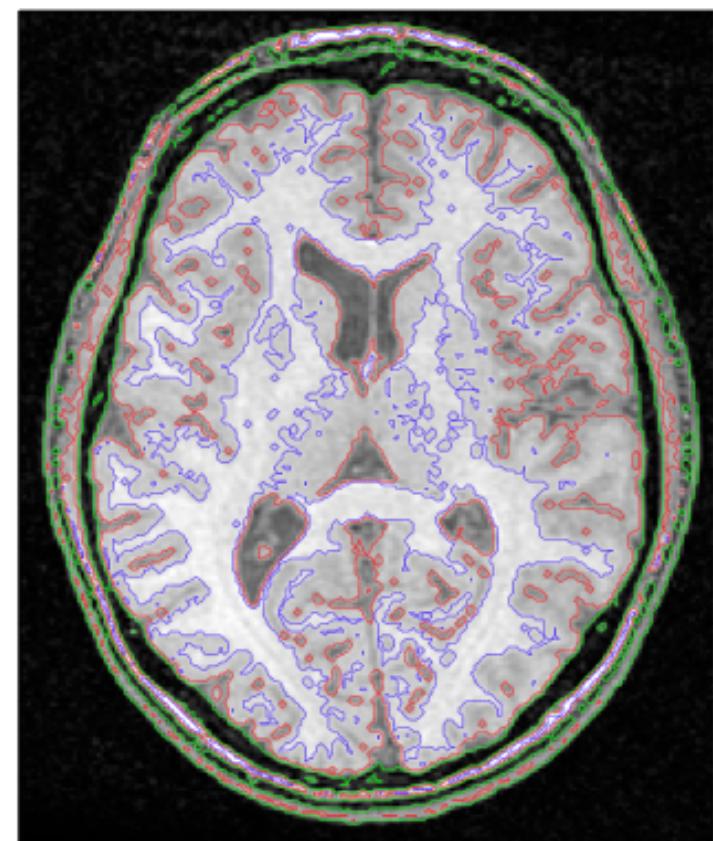
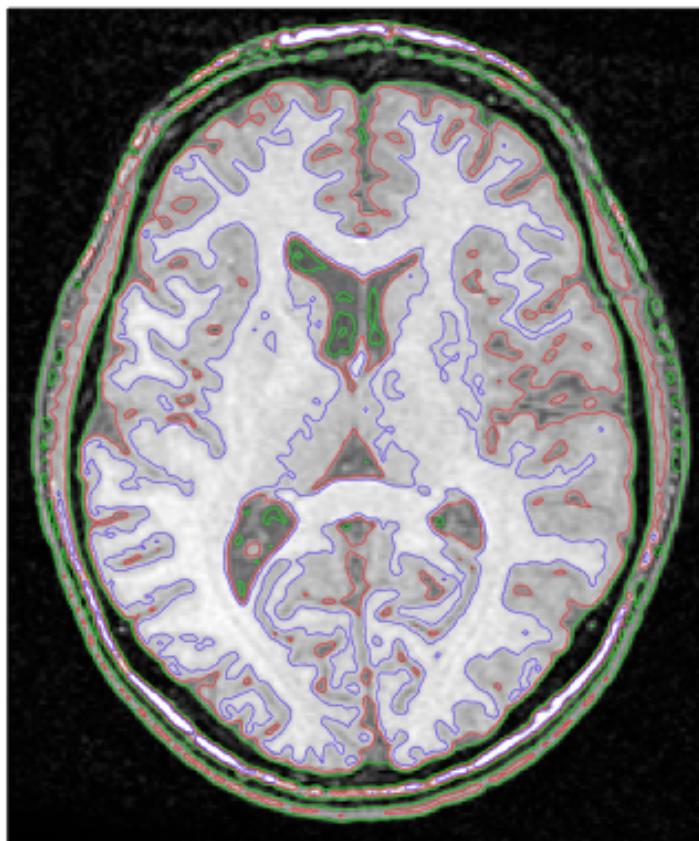


Minimum is obtained
when there is no
intersection:
 $a_1=0, k(S1,S2)=0$

Maximum is obtained
when every points
intersect:
 $a_1=a_2=a_3, k(S1,S2)=1$

$$\kappa(S_1, S_2) = \frac{2a_1}{2a_1 + a_2 + a_3}$$

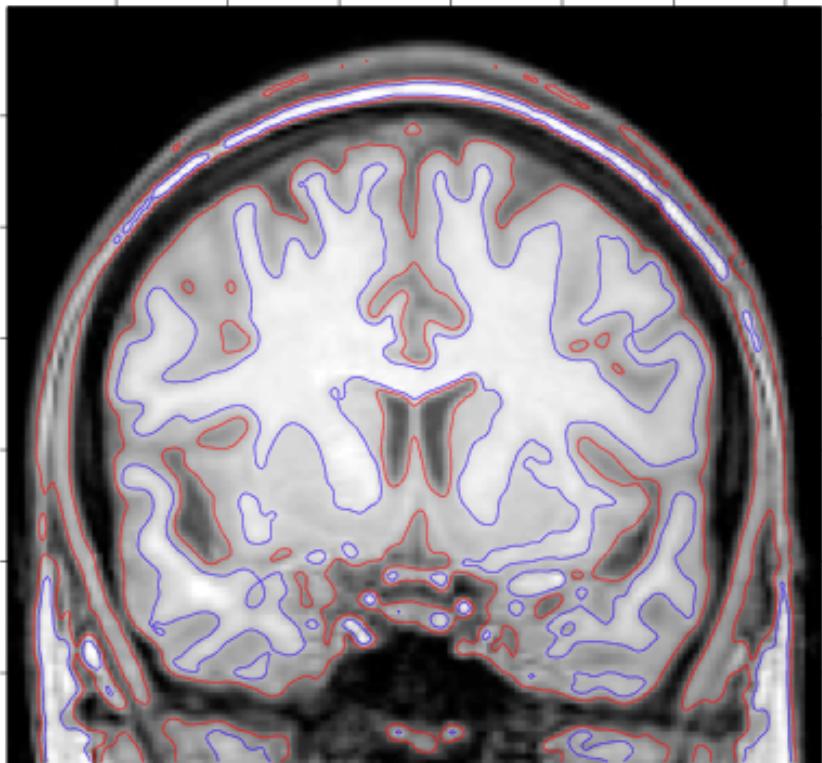
Comparing different segmentation techniques



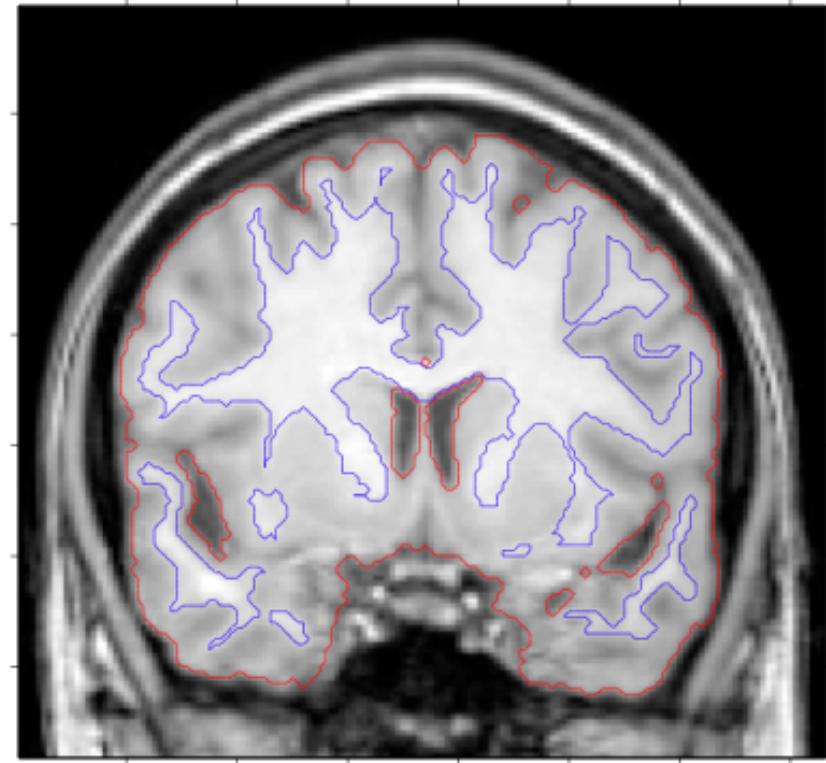
Thin plate spline (TPS)
segmentation

Neural network
segmentation

Automatic vs. manual (gold standard)



TPS segmentation



manual segmentation

Correlation vs. Kappa index

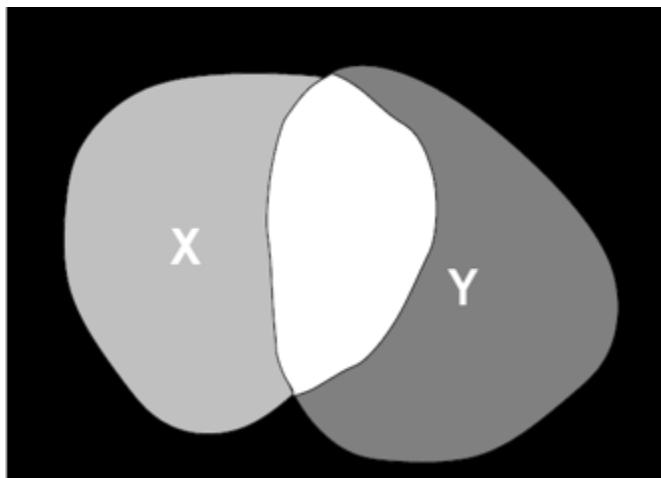
	subject no.	Corr. Coef.		Kappa Index	
		GM	WM	GM	WM
TPS vs Manual	1	0.660	0.827	0.836	0.872
	2	0.702	0.757	0.841	0.827
	3	0.654	0.787	0.811	0.850
	4	0.410	0.678	0.723	0.770
	5	0.612	0.791	0.776	0.838
mean (sd)		0.608(0.115)	0.768(0.056)	0.798(0.049)	0.831(0.038)
SPM vs Manual	1	0.675	0.846	0.883	0.866
	2	0.686	0.839	0.887	0.880
	3	0.637	0.810	0.863	0.842
	4	0.091	0.672	0.679	0.753
	5	0.450	0.803	0.825	0.824
mean (sd)		0.518(0.250)	0.794(0.071)	0.827(0.087)	0.833(0.050)
TPS vs SPM	1	0.806	0.883	0.848	0.900
	2	0.626	0.759	0.794	0.824
	3	0.734	0.822	0.808	0.861
	4	0.426	0.767	0.730	0.793
	5	0.645	0.800	0.785	0.836
mean (sd)		0.647(0.143)	0.806(0.050)	0.793(0.043)	0.843(0.040)

Pearson correlation is *not* a good measure for binary output.
Question. So should Kappa index be used over correlation?

Other set theoretic similarity measures

Williams' index (Bouix et al., 2005)

Hausdorff distance



Jaccard

$$\frac{|X \cap Y|}{|X \cup Y|}$$

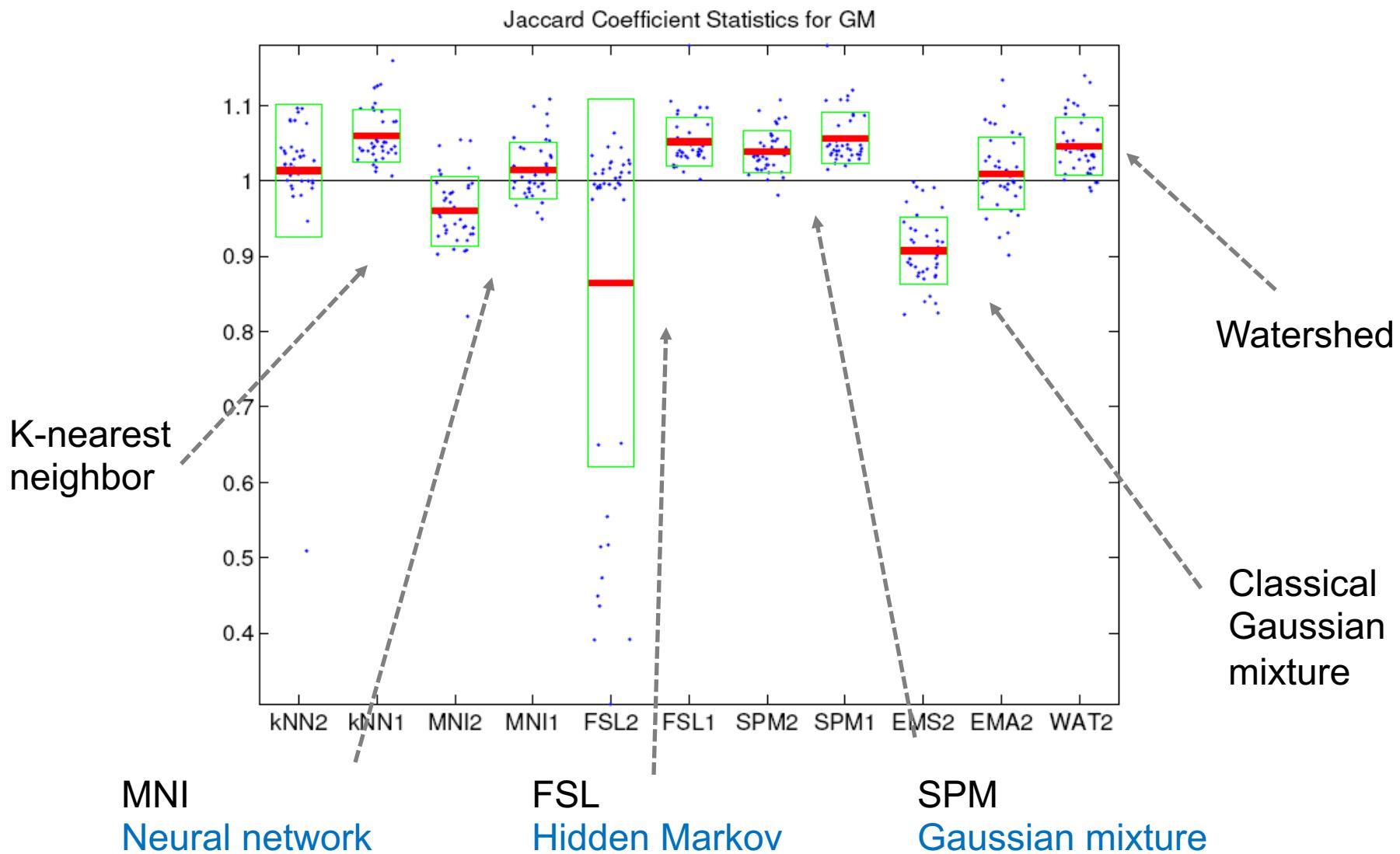
Simple matching:

$$\frac{|X \cap Y| + |\overline{X} \cap \overline{Y}|}{n}$$

Volume similarity:

$$1 - \frac{||X| - |Y||}{|X| + |Y|}$$

Performance analysis of brain image segmentation algorithms before deep learning

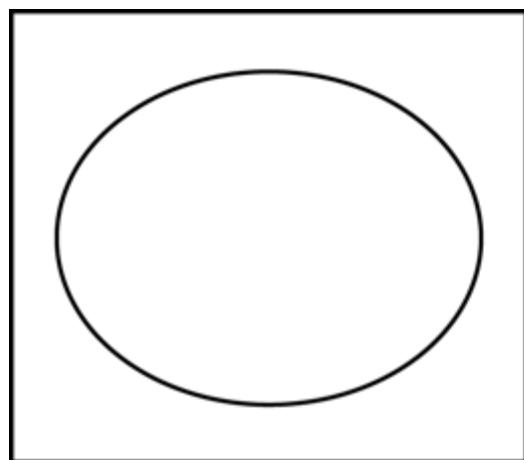


Geometric similarity measures

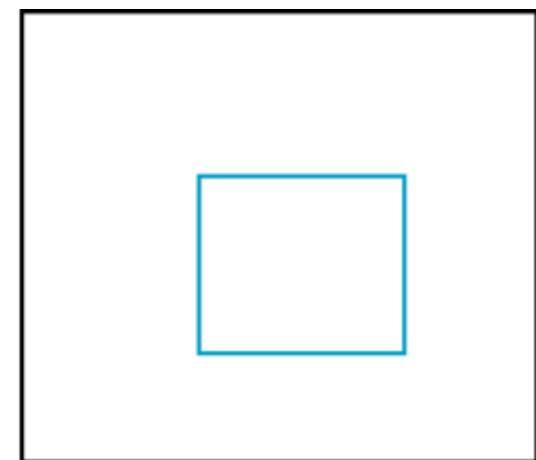
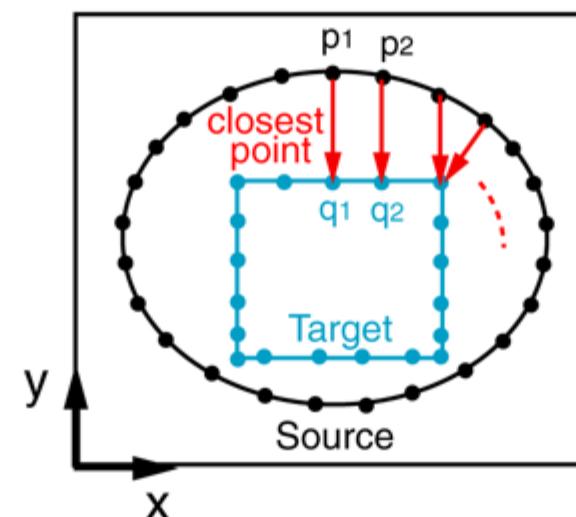
Measure of surface/curve similarity

$$e(X_1, X_2) = \int \min_{p' \in A'} \|X_1(p') - X_2(p)\|^2 dA / \int \|X_2(p)\|^2 dA$$

Staib (1996)



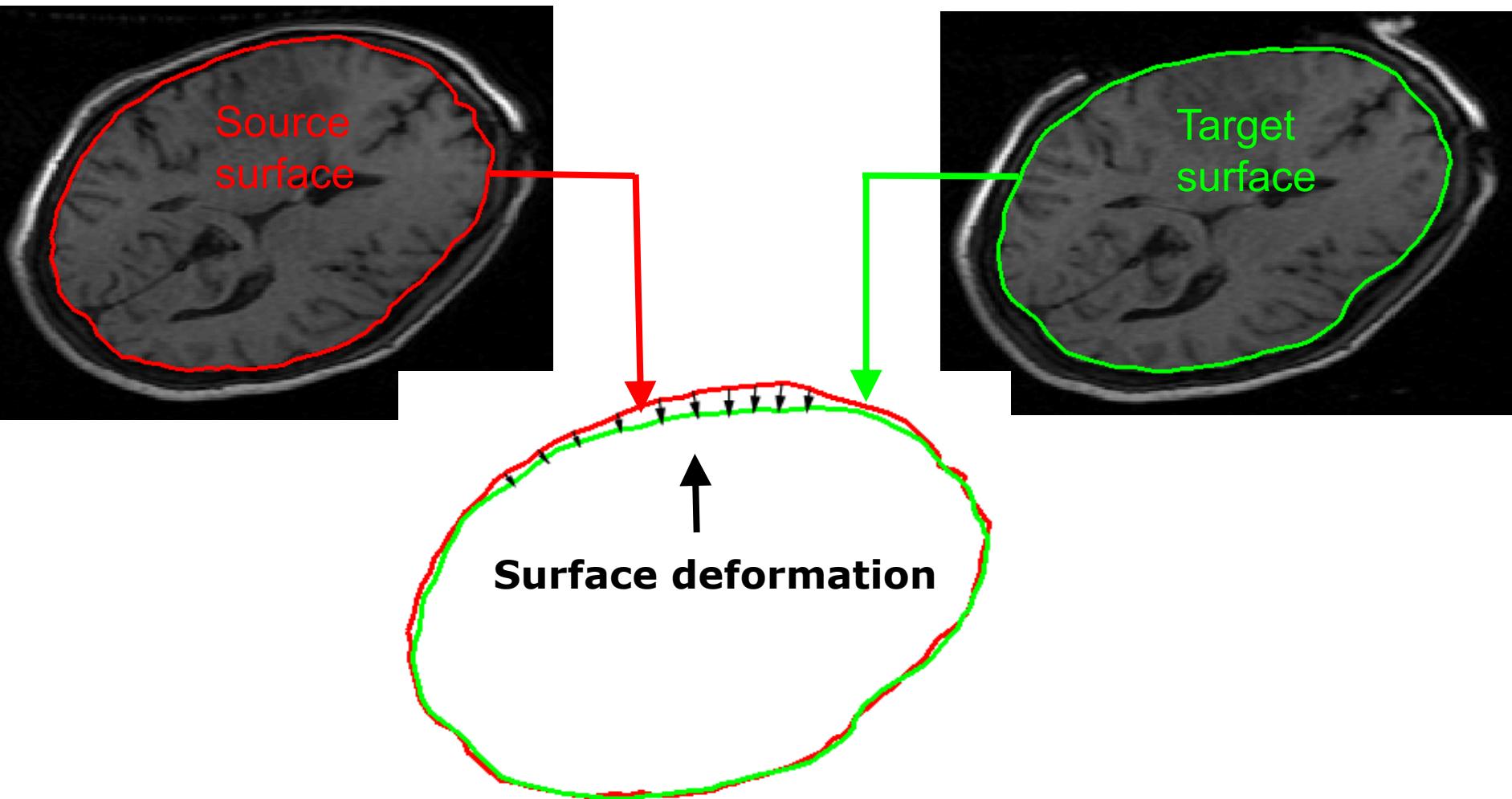
Source



Target

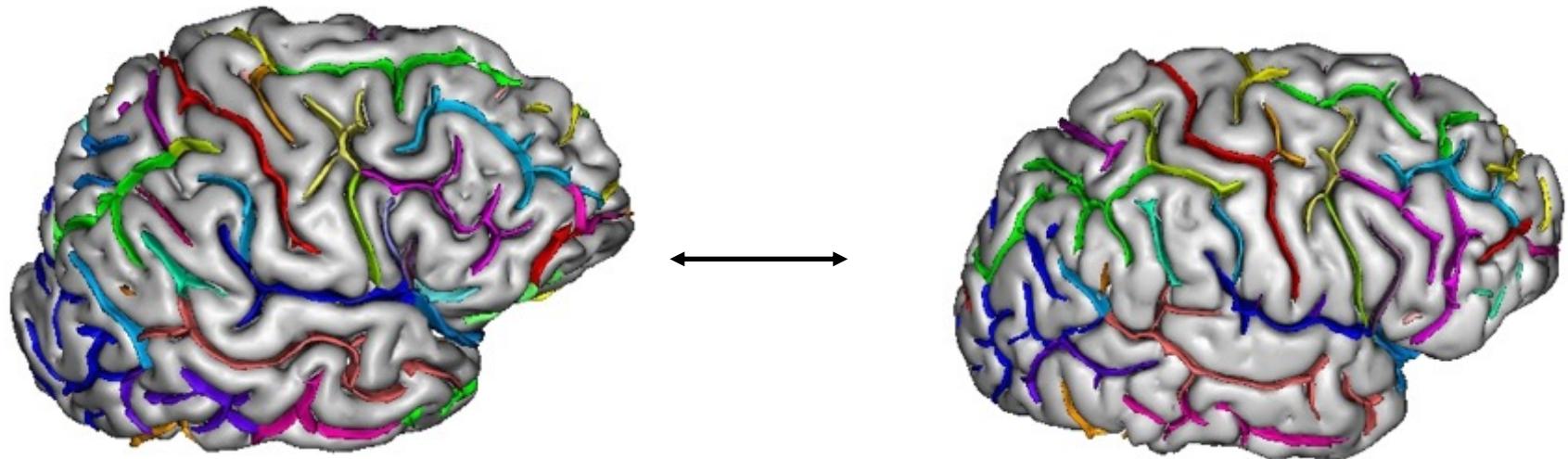
If you use SPHARM/CSR, you don't need this technique (see *required reading*).

Application of geometric similarity measures



Deformable surface modeling: surface-to-surface registration
→ Diffomorphic image registration ([state-of-art](#)): ANTS, SPM

Application of surface similarity measure



Sulcal pattern matching and brain shape analysis

BrainVisa (J.-F. Mangin) output

Information theoretic similarity measures

- Entropy: thermodynamic concept of measuring the amount of chaos in a system
- Shannon (Bell lab, 1948) introduced it as the measure of complexity in information
- Kullback (Kullback & Leibler, 1951) used it in defining the distance between two random variables and distributions

Kullback-Leibler divergence

For probability density f and g , the Kullback-Leibler divergence (KL distance) is defined as

$$H(f, g) = \mathbb{E}_f \ln \frac{f(X)}{g(X)} = \int \log \frac{f(x)}{g(x)} f(x) .dx$$

$$H(f, g) \geq 0$$

Equality is obtained when $H(f, f) = 0$

Smaller $H \rightarrow$ more similar

H is called *relative entropy*.

Kullback-Leibler divergence

Question: Is H symmetric?

$$H(f,g) = H(g,h)?$$

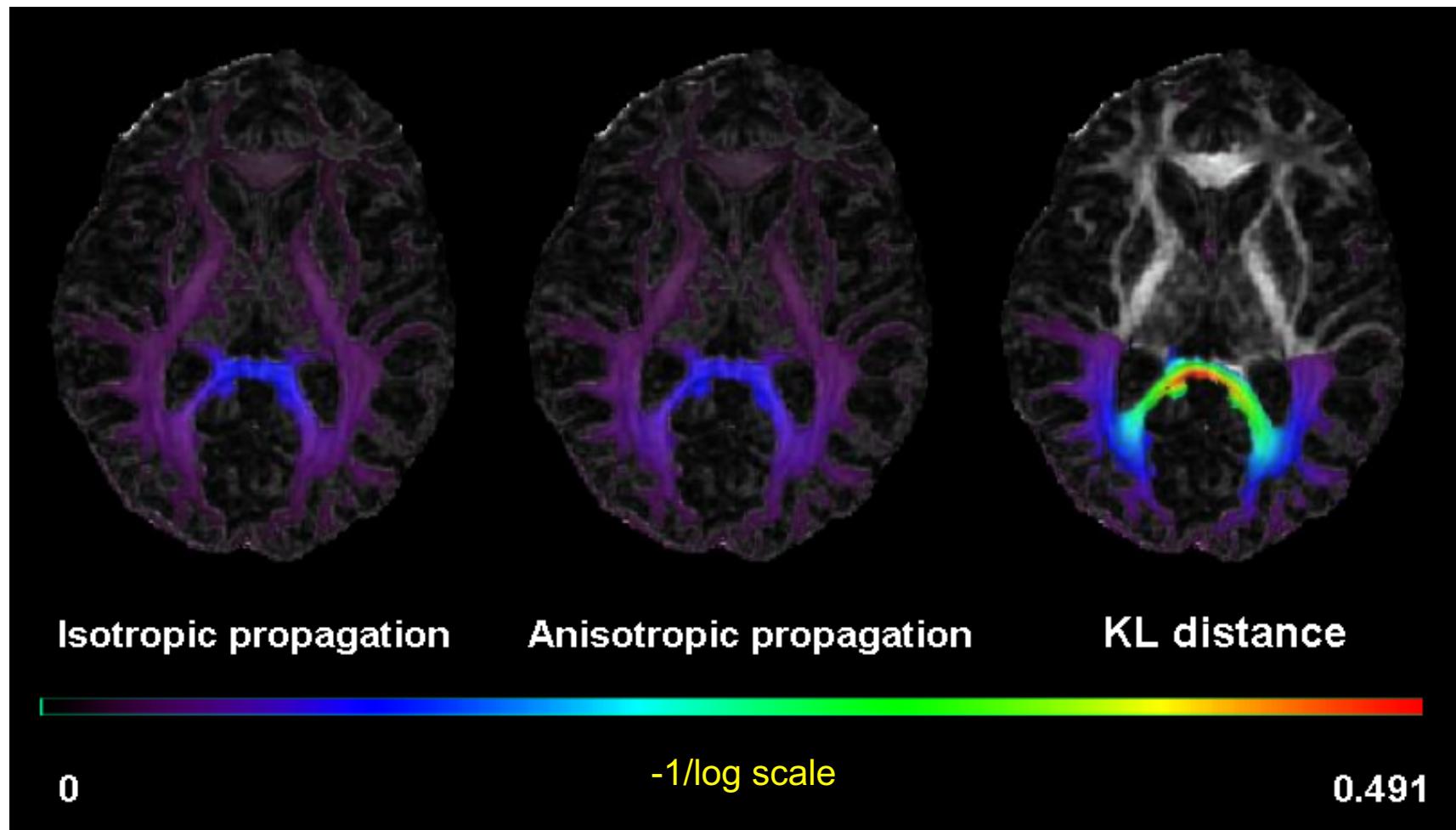
Computing KL-distance

In practice, we estimate the distribution of a random variable via the empirical distribution (histogram).

Discretization:

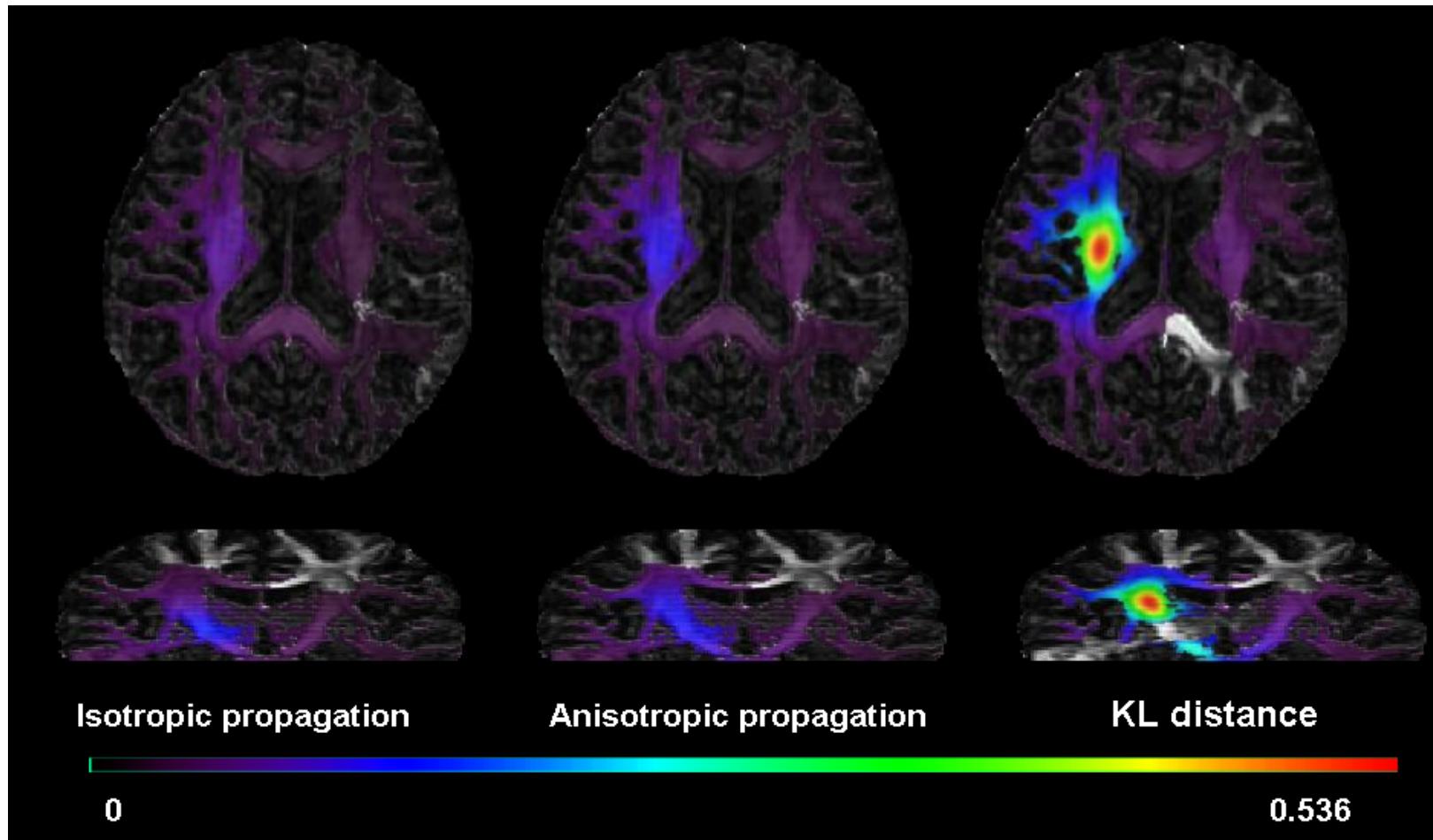
$$\int \ln \frac{f(x)}{g(x)} f(x) dx = \sum_i \ln \frac{f(x_i)}{g(x_i)} f(x_i)$$

DTI connectivity via KL-distance



KL-distance is computed within a small window around each voxel

KL-distance in DTI



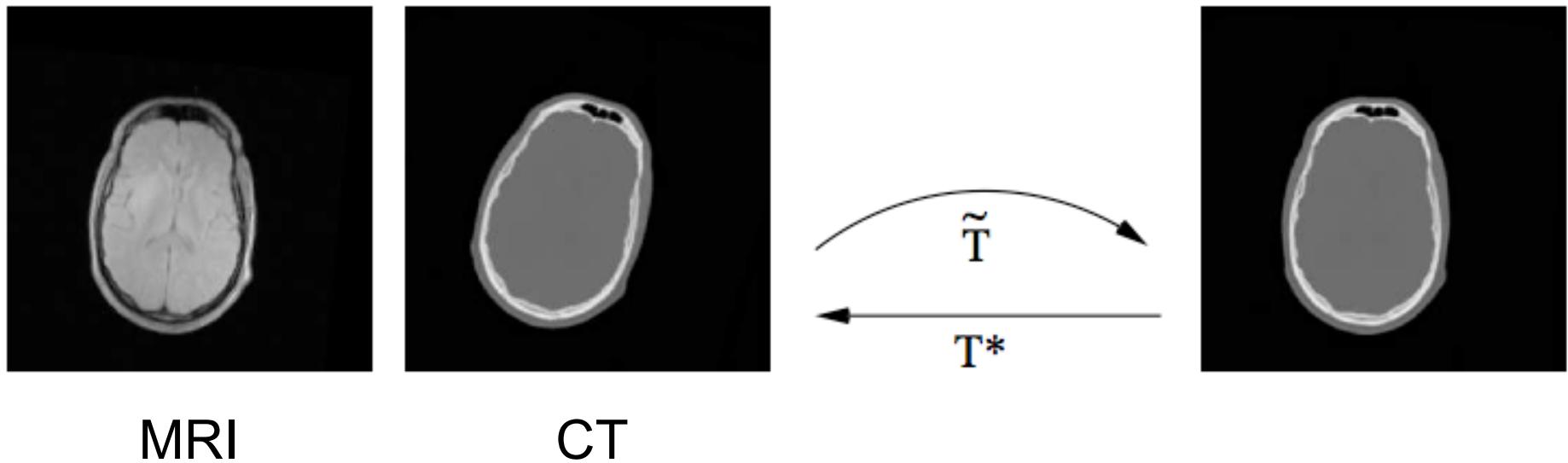
Null model

Alternate model

Significance of connectivity can be also tested.

Other use of KL-distance

Registration cost function for locally varying image intensity histogram



Mutual information

Let f be the joint density of f_1 and f_2 . The mutual information of f_1 and f_2 is defined as

$$I(f_1, f_2) = H(f, f_1 f_2)$$

Mutual information

If $f=f_1f_2$, two random variables are independent. If random variables are independent, $I(f_1, f_2)=0$.

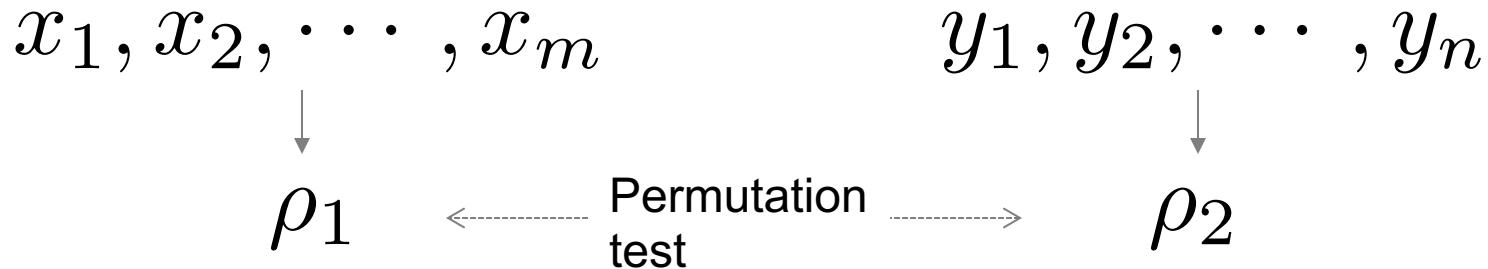
If two random variables are independent, $\text{corr}(f_1, f_2)=0$. So mutual information behaves like correlation.

Unit Outcomes

- 1) Understand the definition of similarity measures.
- 2) Able to compute similarity measures in various data types.

Self assessment questions

- 1) Prove that Pearson correlation is scale and translation invariant
- 2) Suppose there are two groups of data and resulting Pearson correlations.



Perform the permutation test. Compare the result against the parametric approach (Fisher transform + standard normal).