НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Classification between Norm and Pathology Based on Spectral Features of Network Brain Structures

Классификация нормы и патологии на основе спектральных признаков сетевых структур мозга

Юлия Додонова

Classification Based on Neuroimaging Data

Why is it important?

- Classify normal and pathological brain structures

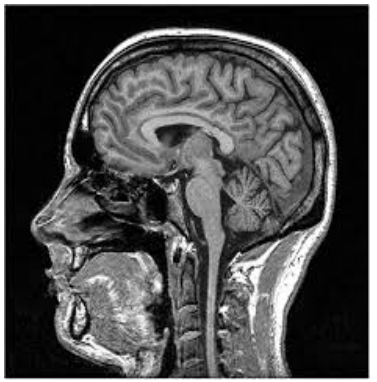
  (Alzheimer's disease, Parkinson disease, Autism Spectral Disorders, etc.)

- Predict treatment outcome

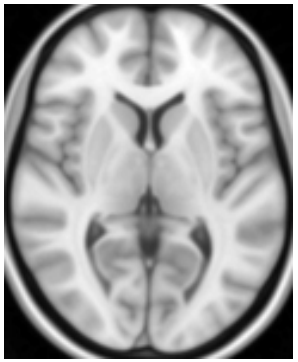
  (Huntington disease, stroke, etc.)

How to tackle it?

- voxel-based
- volume-based
- region-based
- network-based

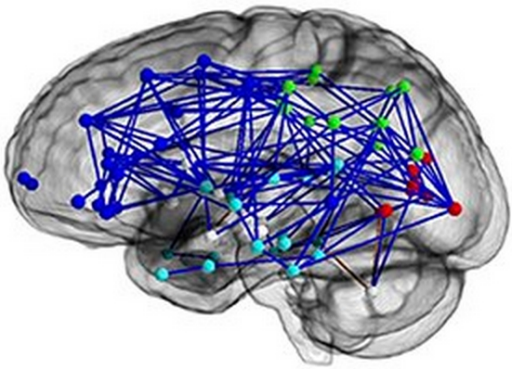


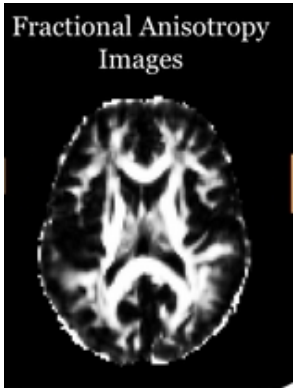
Connectomes: Brain Networks



MRI

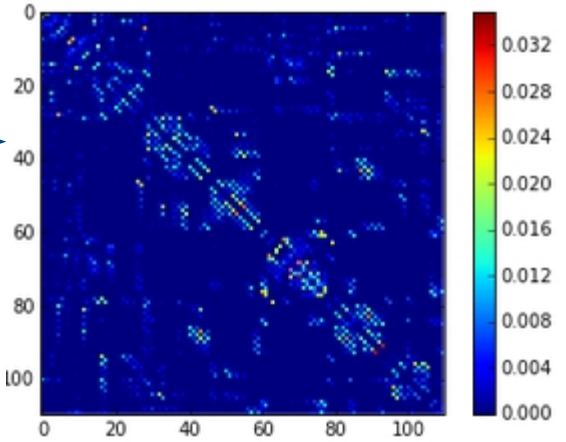
Zone parcellation - vertices





dMRI

Tractography – edges



Adjacency matrix

The term "Connectome" introduced by O.Sporns and P.Hagmann in 2005

Connectomes: general

- Small graphs (~100 vertices)
- Undirected (symmetric adjacency matrices)
- Connected

- Each vertice is uniquely labeled
- A set of labels is the same across networks
- Vertices have 3D coordinates

- Sparse (~10% density)
- Weighted edges (weights are proportional to the number of streamlines between the brain regions)

Connectomes: previous classification studies

How to represent objects to be classified?

- "Bag-of-edges" (huge dimensionality)
- Vectors of local metrics (better, but still a lot)
- Vectors of global metrics (too global?)

  Graph clustering coefficient, graph characteristic path length, small-worldness, modularity, etc.

What are the sample sizes?

~ 50 objects is a typical sample,

~ 100 objects is a large sample

e.g., see a review by Wolfers et al. (2015)

Connectomes: Graph spectra

Spectra of the adjacency matrices

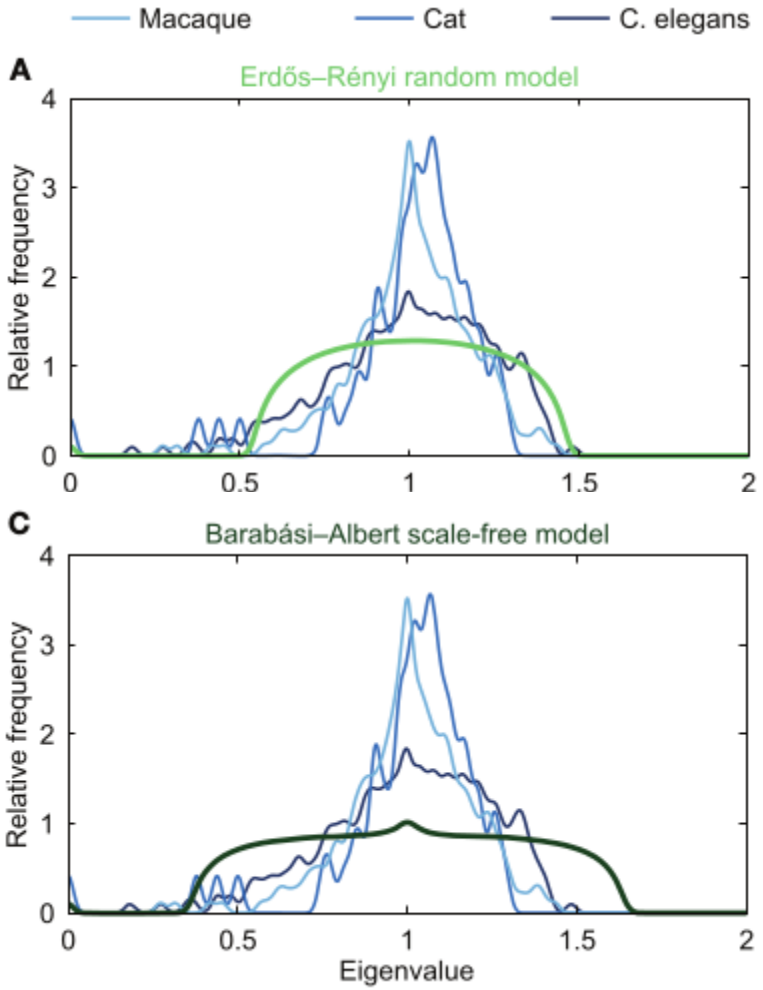
Spectra of the Laplacian matrices

$$L = D - A$$

Spectra of the normalized Laplacians

$$\mathcal{L} = D^{-1/2} L D^{-1/2}$$

The spectra of each of these matrices can be useful in their own way as they capture different graph properties.



Figures from De Lange et al., 2014, distributions of the spectra of the normalized Laplacians of brain networks

Our Study

Develop methods for classification between normal and pathological brain structures based on brain networks

Evaluate how these methods work for two classification tasks:

(1) autism spectrum disorder versus typical development

(2) carriers versus non-carriers of the APOE-4 allele associated with an increased risk of Alzheimer's disease

Our Study: Data

UCLA Autism dataset

- 94 subjects
- 51 ASD subjects (age 13 ± 2.8 years),
  43 TD subjects (age 13.1 ± 2.4 years)
- 264x264 matrices

UCLA APOE-4 dataset

- 55 subjects
- 25 APOE-4 carriers (age 60.8 ± 9.7 years),
  30 APOE-4 non-carriers (age 63.8 ± 8.3 years)
- 110x110 matrices

Data: Baselines

UCLA Autism dataset

- No machine learning baseline provided by the authors of the dataset
- ROC AUC 0.77 obtained in our previous study

UCLA APOE-4 dataset

- No machine learning baseline

In this study, we use the simplest features: vectorized adjacency matrices (“bag of edges”) and the vectors of weighted degrees for the baseline qualification quality

Our Study: Preprocessing

Original weights

- weights proportional to the number of streamlines (~fiber tracts) between brain regions
- each matrix scaled by the sum of its edge weights to enhance between-subject comparison: $a_{ij}^{scaled} = \frac{a_{ij}}{\sum_{i,j} a_{ij}}$

Binary weights

- all non-zero weights set to 1: $a_{ij}^{binarized} = 1 \quad if\, a_{ij} > 0,\ 0\ else$

Weights divided by the squared distances between regions

- used Euclidean distances between regions: $a_{ij}^{weighted} = \frac{a_{ij}}{l_{ij}^2}$
- computed distances from 3D coordinates of zone centers
- each matrix scaled by the sum of its weights afterwards

Classification on the vectors of eigenvalues-1

Features:

- sorted eigenvalues of the graph-related matrices: adjacency, Laplacian and normalized Laplacian
- “bag of edges” and degree vectors for baselines

Classifiers (binary classification):

- Linear classifiers
  - *logistic regression and linear SVM*
  - *with preliminary minmax scaling*
  - *elastic net regularization*
  - *grid search on the two regularization parameters*
- Boosted decision trees
  - *no preliminary scaling*
  - *grid search on the optimal depth of the tree, sub-sample ratio of the features and objects, and regularization coefficients*

Classification on the vectors of eigenvalues - 2

Features	Original weights	Binarized weights	Weights by l^2
Edges	0.525±0.046	0.515±0.000	0.531±0.023
	0.595±0.050	0.510±0.070	0.523±0.075
Degrees	0.539±0.028	0.515±0.000	0.545±0.018
	0.534±0.053	0.515±0.055	**0.754**±0.034
A spectra	0.515±0.000	0.509±0.005	0.516±0.003
	0.687±0.022	0.506±0.052	0.647±0.030
L spectra	0.585±0.043	0.515±0.000	0.534±0.039
	0.501±0.047	0.508±0.031	0.581±0.026
$\mathcal{L}$ spectra	0.514±0.005	0.507±0.040	0.515±0.000
	0.556±0.033	0.504±0.048	0.638±0.037

(a) UCLA Autism dataset

Features	Original weights	Binarized weights	Weights by l^2
Edges	0.552±0.038	0.558±0.028	0.550±0.000
	0.570±0.062	0.563±0.062	0.520±0.072
Degrees	0.550±0.000	0.550±0.000	0.631±0.034
	0.548±0.062	0.545±0.079	**0.704**±0.049
A spectra	0.550±0.000	0.550±0.000	0.550±0.000
	0.717±0.034	**0.780**±0.039	0.533±0.066
L spectra	0.550±0.000	0.550±0.000	0.604±0.011
	0.525±0.065	0.638±0.041	0.512±0.057
$\mathcal{L}$ spectra	0.550±0.000	0.689±0.043	0.592±0.031
	0.584±0.065	0.647±0.039	0.506±0.067

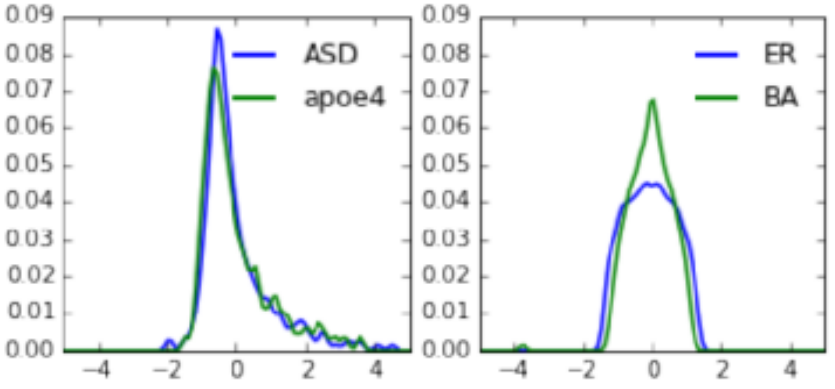
(b) UCLA APOE-4 dataset

Baselines vs. **Vectors of eigenvalues**

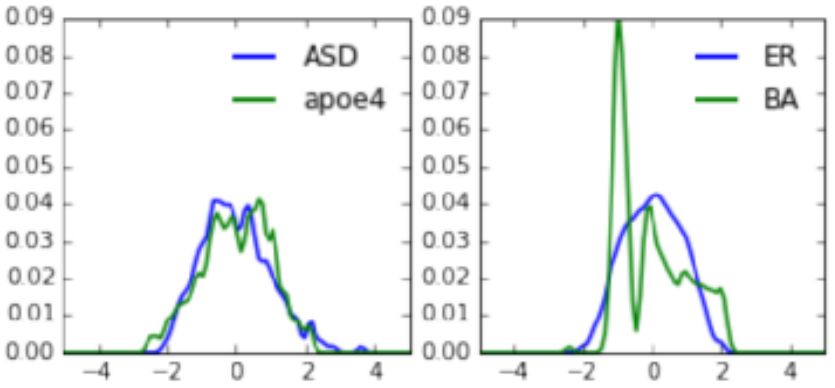
Best classification results: ROC AUC values over 100 different 10-fold splits (mean ± standard deviation).

Within each cell top row: linear models (logistic regression and linear SVM); bottom row: boosted decision trees

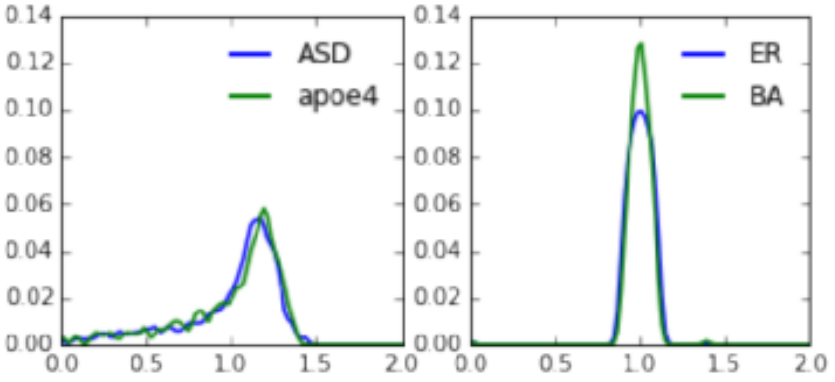
Distributions of eigenvalues



(a) Adjacency matrices



(b) Laplacians



(c) Normalized Laplacians

***Left column**: UCLA Autism (ASD) and UCLA APOE-4 (apoe4)*

***Right column**: random (ER) and preferential attachment (BA) models*

Information divergence based kernels

The Kullback-Leibler kernel:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) log \frac{p(x)}{q(x)} dx$$

$$K_{KL}(p,q) = e^{-\alpha(KL(p||q)+KL(q||p))}$$

The Jensen-Shannon kernel:

$$JS(p||q) = \frac{1}{2}(KL(p||r) + KL(q||r)) \qquad r(x) = \tfrac{1}{2}(p(x) + q(x))$$

$$K_{JS}(p,q) = e^{-\alpha\sqrt{JS(p||q)}}$$

SVM with information divergence based kernels

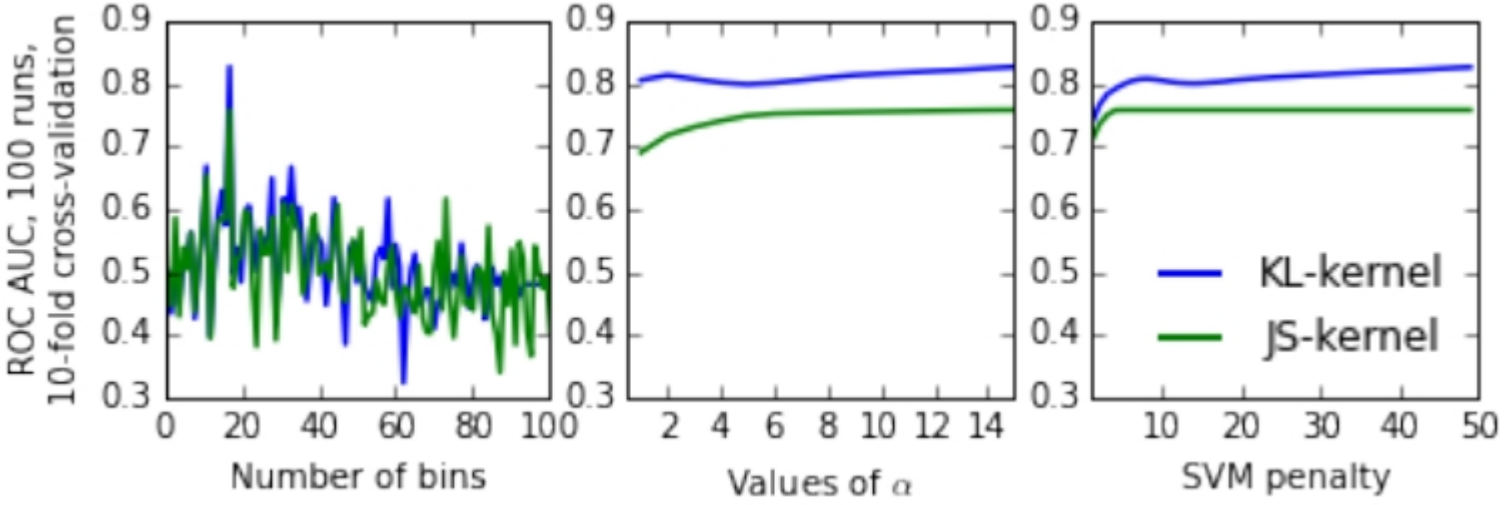
- need explicit density reconstruction: frequencies

- feed SVM with the precomputed KL and JS kernels

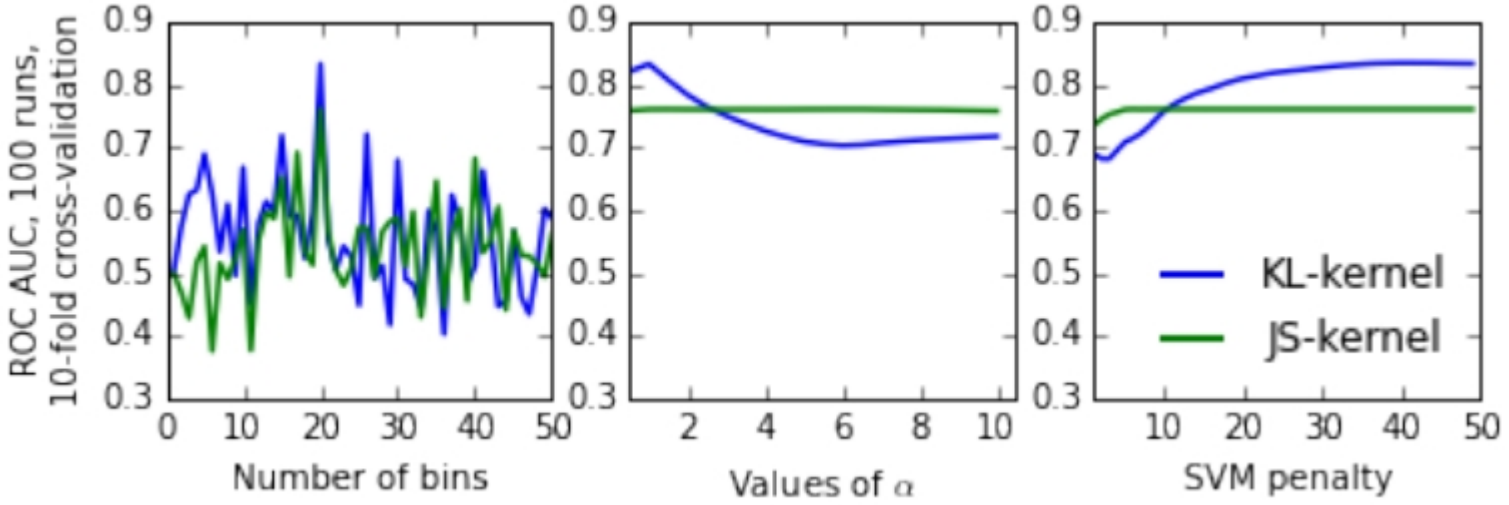
- three parameters:

  (1) number of bins to reconstruct frequencies,
  (2) coefficient of the kernel,
  (3) penalty parameter of the SVM classifier

Information divergence kernels: parameters



UCLA Autism dataset



UCLA APOE-4 dataset

Earth mover's distance based kernel

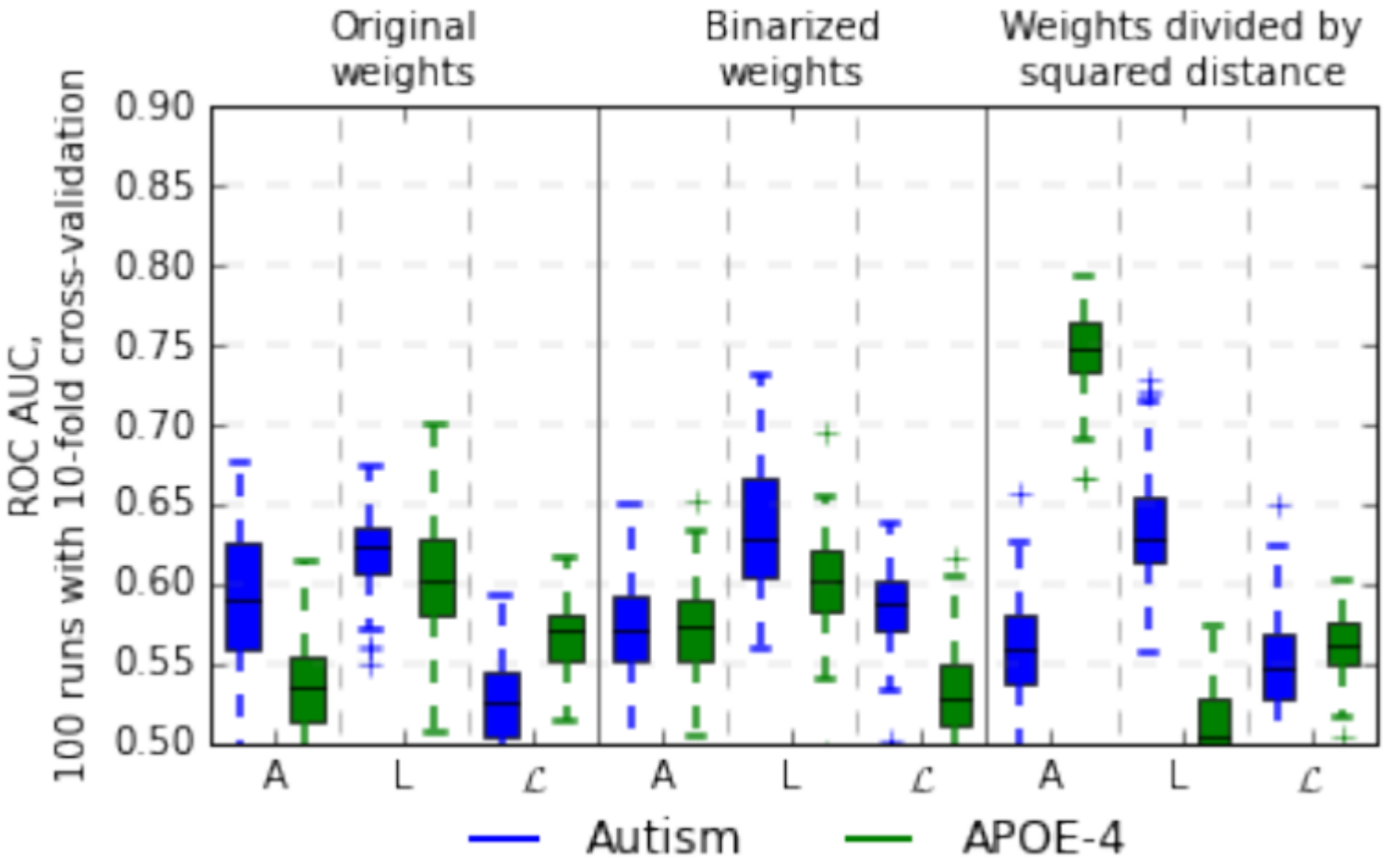
An idea behind this metric:

If each distribution is represented by some amount of dirt, EMD is the minimum cost required to move the dirt of one distribution to produce the other. The cost is the amount of dirt moved times the distance by which it is moved *(we only consider the Euclidean ground distance)*.

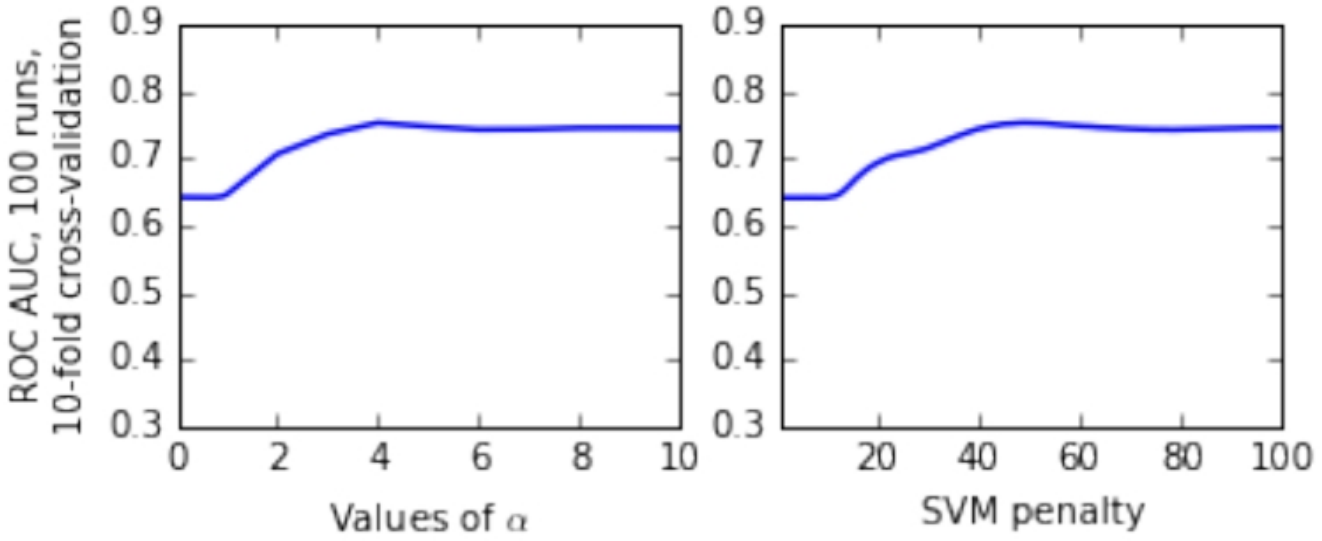
EMD-based kernel:

$$K_{EMD}(P,Q) = e^{-\alpha EMD(P,Q)}$$

Earth mover's distance kernel: results



Earth mover's distance kernel: parameters



UCLA APOE-4 dataset, models on the spectra of the adjacency matrices, weighting by the squared distances

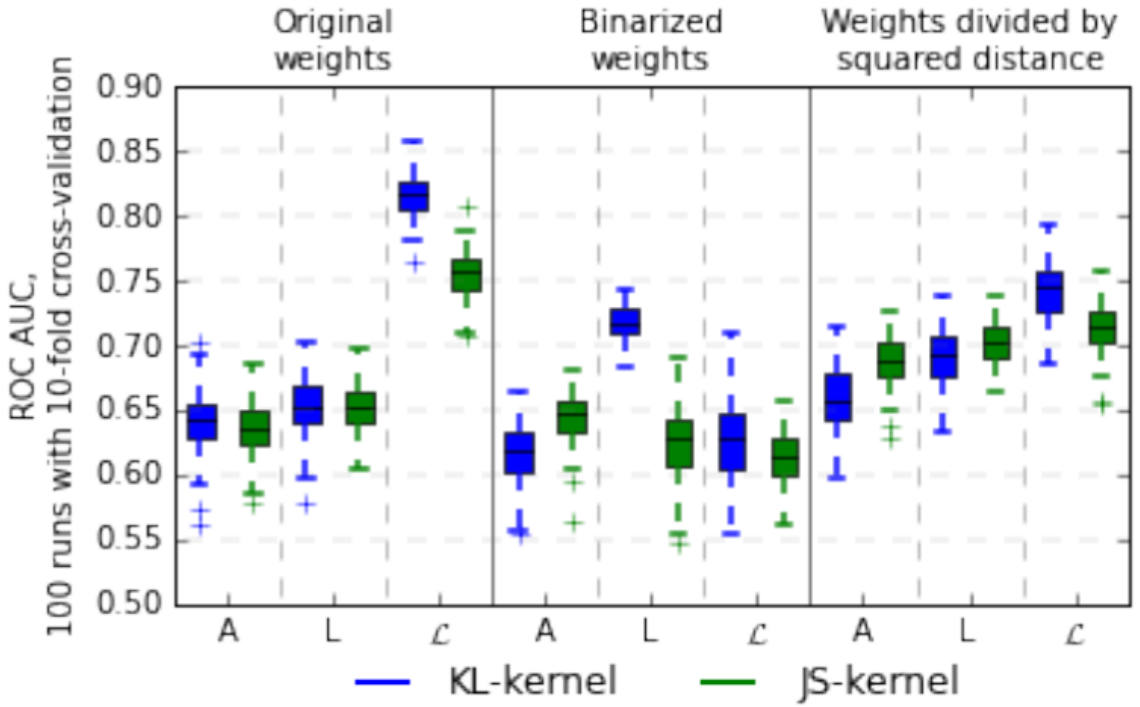
Future work

- Other possible approaches to density reconstruction for information divergrence kernels
- Other approaches to transforming a matrix of pair-wise distances to a positive semi-definite kernel
- Sensitivity of the proposed pipelines to the brain parcellation scheme used to define nodes
- **Verification of the proposed pipelines on the larger datasets with different classification tasks**

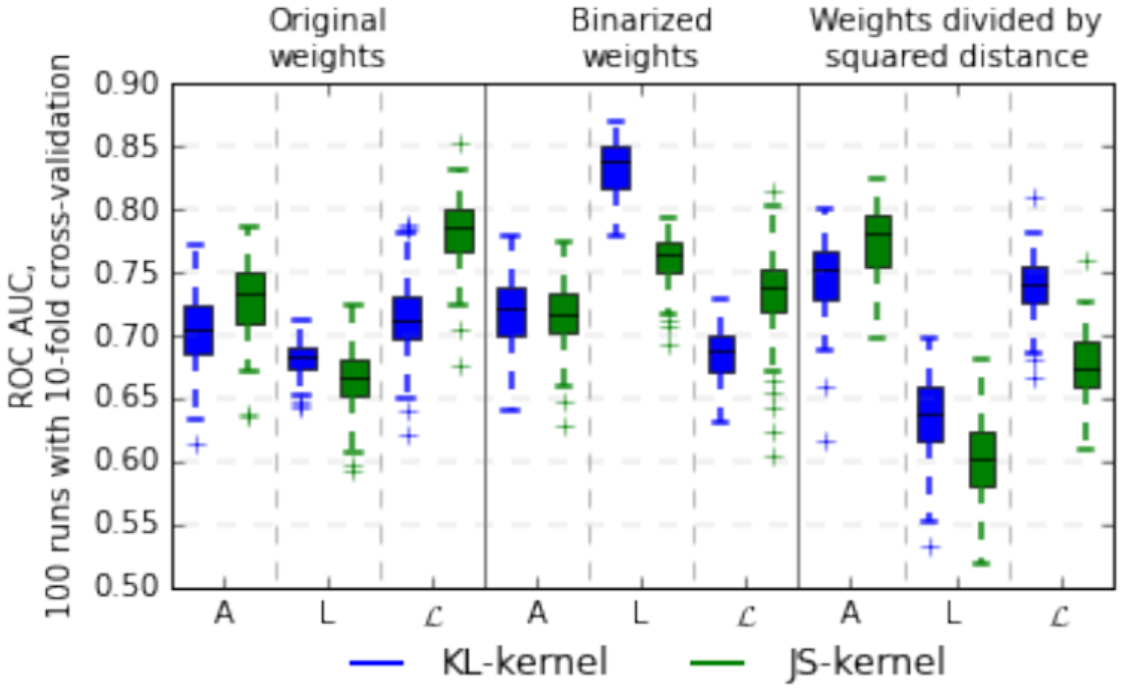
  - we currently work with the much richer datasets on the Parkinson disease and Alzheimer’s disease

Thank you

Information divergence kernels: results



UCLA Autism dataset



UCLA APOE-4 dataset