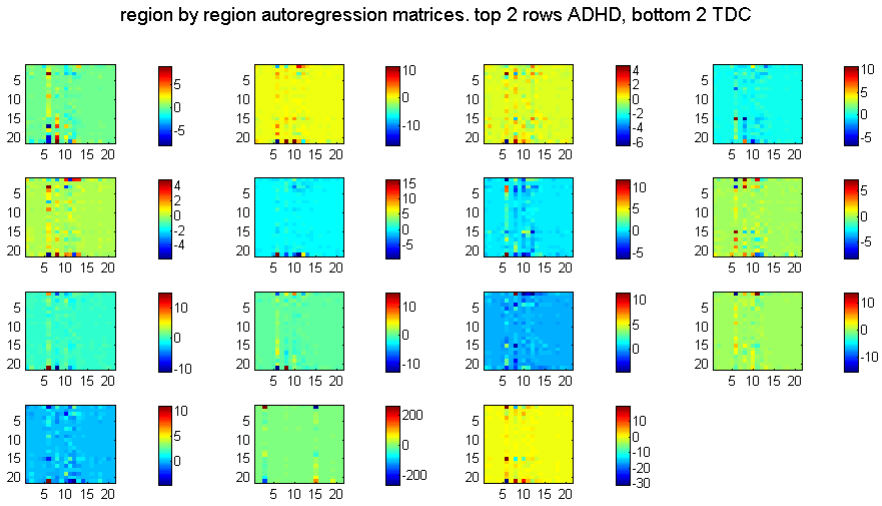
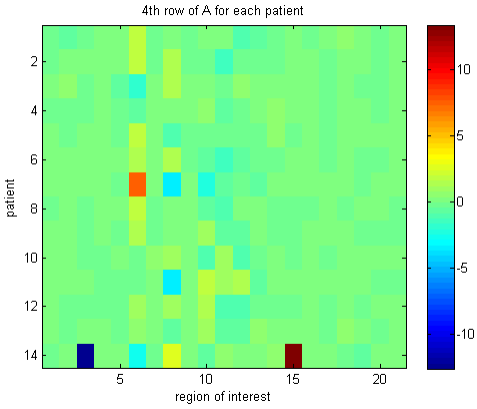
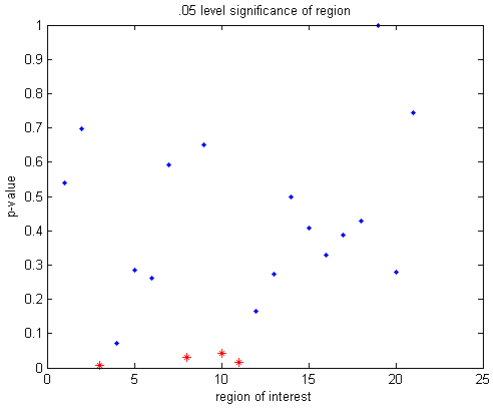
EN.580.694: Statistical Connectomics
Final Project Report

Michelle Chyn · May 14, 2015

Kernel based hypothesis testing on autoregression matrices for comparison of ADHD and TDC patients







Opportunity Currently there are many separate studies on estimating connectomes of specfic disease states or phenotypes such as ADHD, developmental age, IQ, etc.. The next step would be to find voxel connections that are indicators of specific phenotypes and to see what voxelwise connections differ between phenotypes for connectome wide association studies.

Challenge The method used most commonly involved univariate analysis, which was computationally expensive. The novel concept presented by Shezhad et al. in 2014 was to use multivariate distance matrix regression [3], but this method employed the use of a kernel like matrix without taking advantage of the toolbox offered by using kernels. Additionally, many studies do not make use of the time series information within the data that is collected and simply take the average over all time.

Action The Desikan atlas was used to reduce the data size for ease of computation and regions with 0 data over all time were discarded. Instead of averaging over time, matrix autoregression using one time step was employed to create the adjacency matrices. Then, the gaussian kernels of each region of interest versus itself between 2 patients were computed. More specifically the r^{th} row of each A matrix was put into a $n \times r$ matrix, ending up with a matrix whose first 7 rows were ADHC patients with regions affecting the r^{th} region as the columns. The gaussian kernel with $\gamma = 1$ and $\sigma = 1$ was computed between the rows of this matrix [1]. Permutation testing was used by permuting ADHC and TDC labels 10,000 times to create a null distribution

Resolution The first figure above show the adjacency matrices of each patient where each entry $a_{i,j}$ represents the effect of region j on the next time step value of region i. The second figure shows the vectors to be put into the kernel function of a randomly chosen region (4) where $entry_{i,j}$ represents the region j contribution to region 4 in patient i. From these r graphs, the kernel and null distribution was computed, and p-value is shown in the last figure. The stars represents .05 level significance, meaning that 4 regions were significantly different in the contribution of other regions towards itself between ADHD and THC patients.

Future Work Due to the discarding of regions with 0 data, the resulting region labels need to be traced in order to identify what the significant regions are on the atlas. Further comparison of these results to literature [2] are needed to clarify the importance of these results. Also, other phenotypic comparisons would be useful for verifification of this method.

References

[1] Harchaoui, Z., Bach, F., Cappe, O., & Moulines, E. (2013). Kernel-Based Methods for Hypothesis Testing: A Unified View. *IEEE Signal Processing Magazine*, 87-97.

[2] Neurosynth. (n.d.). Retrieved April 1, 2015, from http://neurosynth.org/

[3] Shehzad, Z., Kelly, C., Reiss, P., Craddock, R., Emerson, J., Mcmahon, K., . . . Milham, M. (2014). A multivariate distance-based analytic framework for connectome-wide association studies. *NeuroImage*, 74-94.