Homework 5

Background
Listen to the lecture on cross-validated distances. For additional background reading, have a look at the Walther et al. (2016) paper that is posted on the website.

1. Why is the simple Euclidean distance on noisy data a biased estimate of the true distance between two activity patterns?

Data

The data for this homework is dataset_5. It contains activation estimates from the motor cortex of 12 subjects during the movement of the contralateral(CONTRA) fingers. We also give the activation patterns for the ipsilateteral fingers (IPSI), but you don't need them for this homework. Each variable is a cell array which contains 3 dimensional "matrix" (a tensor) of size numFing x numVox x numRuns.

For example
`CONTRA{s}(f,p,r):`
Contains the data from the s^{th} subject (12 total) for the f^{th} finger (1-5), the p^{th} voxel and the r^{th} run.

Note: The activation estimates are already spatially pre-whitened, so when you calculate the Euclidean distance on the pre-whitened patterns, you actually calculate the Mahalanobis distance.

2. Do pair-wise classification between activity patterns using the provided pre-whitened estimates. That is, you need to do 10 total classifications, each finger against each finger. Report the classification accuracies for each pair. Visualize the dissimilarity measure (% accuracy) in a 5x5 Representational Dissimilarity Matrix (RDM).

3. Average the activity pattern for each finger across runs. Calculate the 10 pairwise ***squared*** Euclidean distances between the mean activity patterns. Again, show your results using an RDM.
MATLAB Tip: The functions pdist and squareform can make your life much easier.
In a scatterplot, plot the squared Euclidean distances against % accuracy. Use different colours / symbols for the different subjects. What do you observe?

4. Calculate the cross-validated squared Euclidean distance between each of the 10 finger pairs. Report your results in form of a RDM matrix. In a scatterplot, plot the cross validated squared Euclidean distances against the squared Euclidean distance. Use different colours / symbols for the different subjects. What do you observe?

5. How stable is the RDM estimate across subjects? Calculate the correlation between the the vectors of 10 dissimilarities (%accuracies, sq. distance, sq. cross validated distance) for each pair of subjects. Report the average inter-subject correlation for each measure. Which measure yields the highest average correlation?