Note: I'm not sure at all what's the proper way of compressing the 4-classification into a single statistic. Ask Martin C..

I've been working (A LOT) on implementing my own MVPA and RSA toolboxes. The MVPA Princeton toolbox is doing many things I don't understand. Between others, it has a weird way of doing the searchlight.

Hamed's toolbox is also not very flexible.

Chris doesn't like the RSA approach, so I've decided to give up on convincing him and get MVPA working first – but might take back RSA later.

Between others, I've improved the preprocess script. Now I can use it to normalise (or do the inverse transformation!) any file. I'm using this to

1) Move masks from MNI to subject space, because the MVPA is better if done on the 'realign' data

2) I will also use this for searchlight, to normalise the t-statistics map from subject space to MNI space (and then i can also smooth it).

I think I have a better understanding of using regressions for classification. It's wrong to use them if the mask has more voxels than numbers of trials in the training set. However, this doesn't mean that there's not enough information to learn. A regression is trying to learn something else (more) than a classification.

Instead, I've created a classifier that learns the average pattern (i.e. prototype) for each condition. Then it measures the distance from each pattern in evaluation to the prototypes and picks the closest. This distance can be weighted by the standard deviation (as SDT suggests).

With the MVPA code and this classifier, I seem to be able to decode significantly (up to t = +3.90; p<0.001) on switch vs stay with the univariate effect mask.

The problem is that I'm also able to decode from anatomical masks of any lobes (from xjview), whether parietal, frontal, temporal or occipital. I don't know if there's enough information there, or if there's something wrong with my code.

The second seems more likely.

----

This is a summary e-mail I've sent to myself.

this is a summary (i'm going to stop working on subway today to do some reading and focus on curriculum).

**yesterday**

from the chat with chris yesterday, i explained that we generally don't have enough trials to use a regression as a classifier. that's because the number of voxels (i.e. weights to estimate) can't be bigger than the number of trials.  
  
also, i was running the GLM with one regressor for each event. this is because

1) the result of the MVPA won't depend on the GLM used

2) then i don't need to re-run the GLM for every different regressor i want to try to decode

we decided that this method probably would decrease the statistical power and that it was worth trying to

1) running a normal GLM, where each condition has one regressor

2) alternatively, average the patterns of the GLM by session/condition before running the MVPA

these are two different problems (having enough trials to train the classifier; and which GLM to run).

**today**

for the problem of the decoder, i think i've found a solution (at least a temporary one). it's true that we cannot run a regression with a big mask, but that doesn't mean that there's not enough information for a classifier to learn. i've created a very simple decoder based on SDT (it takes the mean of each condition, as if it was a prototype, and makes a decision based on distance to each protoype). this seems to work, at least for decoding switch vs stay :  
GLOBAL "switch" significant  
d-prime        = +0.29 ± 0.08

p-value        = 0.002            
t-statistic    = +3.36

this statistic is obtained by applying a mask extracted from the univariate effect (somewhere between parietal and frontal).  
however, we also find a significant decoding from anatomical frontal, parietal and occipital, and temporal lobes. i wonder if there's something wrong with my code...  
  
i'm still puzzled about papers that say something like this :  
«A penalized logistic regression classifier with L2-norm regularization was implemented using the Princeton Multi-Voxel Pattern Analysis Toolbox ([http://www.pni.princeton.edu/mvpa](http://www.pni.princeton.edu/mvpa" \t "_blank)). Before classification, the top 1000 voxels from the whole brain that best differentiated faces versus objects were selected by running an ANOVA on the univariate data.»  
i think that using a regression as a classifier only makes sense with a small mask (e.g. using searchlight). i'll should do some research about this.  
  
  
as for the other problem (the GLM), after the meeting i realised that the other thing we wanted to try (to run the GLM with less betas/regressors) is actually going to make things worse for a regression (we would have even less samples for our regression). i've tried the second approach ( average the patterns of the GLM by session/condition) but the statistics of the decoding become weaker (around +2.30) (actually less: +2.19)- despite the univariate effect (averaged across subjects with raw data) is the same.