

MVPA BY CROSS-VALIDATED MANOVA

This is an implementation for Matlab of the method introduced by Carsten Allefeld and John-Dylan Haynes, ‘Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA’, *NeuroImage*, 89:345–357, 2014.

PREREQUISITES

Cross-validated MANOVA is based on a (‘first-level’) multivariate General Linear Model. This model has to be specified and estimated in SPM before using these functions.

Estimation of the model is necessary in order to access SPM’s estimates of various fMRI data properties, especially the temporal correlation of the errors. The functions use the generated `SPM.mat` file and the data files referenced therein, as well as the analysis brain mask image. Other files generated during estimation can be deleted.

SEARCHLIGHT ANALYSIS

The main interface is given by the function

```
cvManovaSearchlight(dirName, slRadius, Cs, permute)
```

which computes the cross-validated MANOVA on a searchlight. `dirName` is the name of the directory where the `SPM.mat` file is located, `slRadius` is the radius of the searchlight in voxels, and `Cs` is a cell array whose elements are contrast matrices (see below). `permute` specifies whether permutation values should be computed and defaults to false.

The searchlight radius is interpreted such that every voxel is included for which the distance from the center voxel is *smaller than or equal to* the radius. This means that 0 leads to a searchlight size of 1 voxel, 1 to 7 voxels, 2 to 33 voxels, and so on. This definition may differ from the one used in other implementations of MVPA algorithms and in publications. Note that it is possible to use fractional values for the searchlight radius. For a table of searchlight radii leading to different searchlight sizes, run `slSize`.

The result of the analysis are estimates of a multivariate measure of effect size, the pattern discriminability D , which is intended as a drop-in replacement for the conventional measure of classification accuracy. Statistical parametric maps of D are written to images with filenames of the form

```
spmD_C####_P####.nii
```

enumerating all contrasts and permutations, in the same directory as the SPM.mat file.

While D is the main measure of multivariate effect size of cross-validated MANOVA and the statistic whose values should be reported, for the purpose of statistical inference the *standardized* pattern distinctness may be better suited (see Eq. 17 in the paper). Statistical parametric maps of standardized D are written to images with filenames of the form

```
spmDs_C####_P####.nii
```

Additionally, an image of the numbers of voxels contained in each searchlight is written to VPSL.nii, and the analysis parameters are saved to cmsParameters.mat.

REGION-OF-INTEREST ANALYSIS

The paper describes cross-validated MANOVA only for searchlight analyses. This package contains additional code for ROI analyses. The function

```
[D, p] = cvManovaRegion(dirName, regions, Cs, permute)
```

performs ROI-based analysis on a set of voxels, specified by the parameter `regions` in the form of a logical volume or an image filename, or a cell array to process several ROIs at once.

Since the result of ROI analysis is not an image but a scalar, it is directly returned by the function.

Note that the order of parameters of `cvManovaRegion` has changed with respect to version 2. Please adjust your code accordingly!

CONTRASTS

In cross-validated MANOVA, effects of interest are specified in the form of contrast vectors or matrices, in the same way as for univariate analysis. Simple ('t-like') contrasts are specified as a *column vector*, complex ('F-like') contrasts as a matrix of several columns. Please note that this is the transpose of the format used in the SPM user interface, but identical to SPM's internal format.

The rows of a contrast matrix correspond to the model regressors for each session *separately*, i.e. other than in SPM the contrast should not be explicitly replicated for several sessions. Instead, the program performs the replication internally, assuming that (at least the leading) regressors for each session model the same effects. If there are fewer rows in a contrast matrix than there are regressors for a session, the

matrix is zero-padded. This makes it easy to ignore regressors that may be present in only some sessions or subjects, e.g. modeling error trials, as long as they are trailing.

To ease the specification of contrasts, the utility function `contrasts` can be used to generate contrast matrices for all main effects and interactions of a factorial design, in a form suitable for use with `cvManovaSearchlight`. For example, `Cs = contrasts([2 3])` is equivalent to

```
Cs = { [ 1  1  1 -1 -1 -1] '
        [ 1 -1  0  1 -1  0 ; 0  1 -1  0  1 -1] '
        [ 1 -1  0 -1  1  0 ; 0  1 -1  0 -1  1] ' };
```

describing the two main effects and the interaction.

Note that the resulting contrasts may have to be modified in the case of including HRF derivatives or using FIR models, or if a factor is nested in another one.

REMARKS

- The estimation of D is based on the GLM residuals and therefore depends on a properly specified model. That means that all effects that are known to systematically occur should be included in the model. Because sub-effects can be selected through the mechanism of contrasts, it is neither necessary nor advisable to use different GLMs as the basis of different MVPA analyses.
- The fMRI model specification must include the modeling of temporal autocorrelations in order to correctly estimate the pattern distinctness. For this, the option ‘serial correlations’ in SPM has to be kept at the default value $AR(1)$.
- The functions are optimized for the computation of several contrasts (and permutations) in one run. One call of `cvManovaSearchlight` with several contrasts will take substantially less time than several calls for each contrast separately.
- For computational efficiency, the functions read the complete data set into memory. The analysis should therefore be run on a computer with a sufficient amount of main memory, and using other memory-intensive programs at the same time should be avoided. Peak memory usage is about twice the amount to load the data set, which is (number of in-mask voxels) \times (number of scans) \times 8 bytes.
- `cvManovaSearchlight` contains a checkpointing mechanism. If the computation is interrupted for reasons other than an internal error and then restarted with the same parameters, it picks up at the last checkpoint. Intermediate results are stored in a file `cmsCheckpoint####.mat` in the same directory as `SPM.mat`, where `####` is a hash encoding the parameters. The file is deleted after the computation finishes successfully.

EXAMPLE AND TEST SCRIPT

The subdirectory `cvManovaTest` contains a script `cvManovaTest` that analyses the data of Haxby et al. (2001), both as a test of the implementation and as an example for how to use it.

REGULARIZATION

For very large searchlight sizes or for a large ROI, it is possible that there are so many voxels that adequate estimation of the error covariance matrix is no longer possible. The resulting numerical instability of the method may be remedied by using regularization of the matrix. This implementation contains the option to apply shrinkage towards the covariance matrix diagonal, by supplying the shrinkage parameter `lambda` as an additional parameter to `cvManovaSearchlight` or `cvManovaRegion`.

Note however that with regularization, D is no longer an unbiased estimator of the true pattern distinctness (unless it is zero). It is therefore recommended to avoid regularization, and rather reduce the number of voxels. The recommended searchlight radius for cross-validated MANOVA is 3, leading to a searchlight of 123 voxels. If regularization is used, the parameter should be kept very small, e.g. 0.001.

The implementation contains a hard-coded limit on the number of voxels within a searchlight or ROI regardless of regularization, of 90% of the available error degrees of freedom.

Note that previous experimental code to estimate the optimal shrinkage parameter based on the method of Schäfer and Strimmer (2005) in version 2 has been removed, because it proved to be unreliable.

NEGATIVE PATTERN DISTINCTNESS?

Even though pattern distinctness D is a ratio of explained multivariate variance, or a generalized squared distance, the values provided by cross-validated MANOVA may be negative. That raises the question how such values should be interpreted.

The simple answer is: Negative values do not have an interpretation per se, and they can never be significantly above zero, so there is no problem for reporting.

The longer answer is that the values produced by the algorithm are only estimates of the true pattern distinctness, and these estimates randomly vary around the true value because of a finite amount of data. The true pattern distinctness can never be below zero. However, the estimator was designed to be unbiased (correct on average), and that implies that if the true value is zero or close to it, estimates vary around zero, and therefore about half of them have to be below zero.

This has an exact analogue in the case of cross-validated classification accuracy. The true accuracy can never be below chance level, but estimated accuracies can be.

In some cases, the estimated value of pattern distinctness strongly indicates that the true value is below zero, too. This is most likely the result of a violation of the assumption underlying cross-validation, that the different parts of the data (sessions) are generated in exactly the same way. This assumption may not hold if there are unmodelled confounds in the data, or problems with the design itself. *Strongly* negative estimated values of pattern distinctness therefore suggest that you should recheck your design matrix, or the design itself. Again, the same problem will most likely occur with cross-validated classification accuracy computed from the same data.

Feel free to contact me with questions and comments. Bug reports and feature requests can also submitted via the GitHub issue tracker.

This software was developed with SPM8 and SPM12 under Matlab 7.11–8.5 (R2010b–R2015a), but later versions should work, too. It is copyrighted © 2013–2016 by Carsten Allefeld and released under the terms of the GNU General Public License, version 3 or later.

This file is part of v3 of cvmanova, see
<https://github.com/allefeld/cvmanova/releases>