PyMVPA

http://www.pymvpa.org

A Python Toolbox for Machine Learning-based Data Analysis

Michael Hanke[1], Yaroslav O. Halchenko[2], Per B. Sederberg[3], Stephen José Hanson[2], James V. Haxby[4], Stefan Pollmann[1]

1) Dept. of Experimental Psychology, University of Magdeburg, Germany
2) Psychology Dept., Rutgers-Newark University, NJ, USA
3) Dept. of Psychology, Princeton University, NJ, USA
4) Dept. of Psychological and Brain Sciences, Dartmouth College, NH, USA

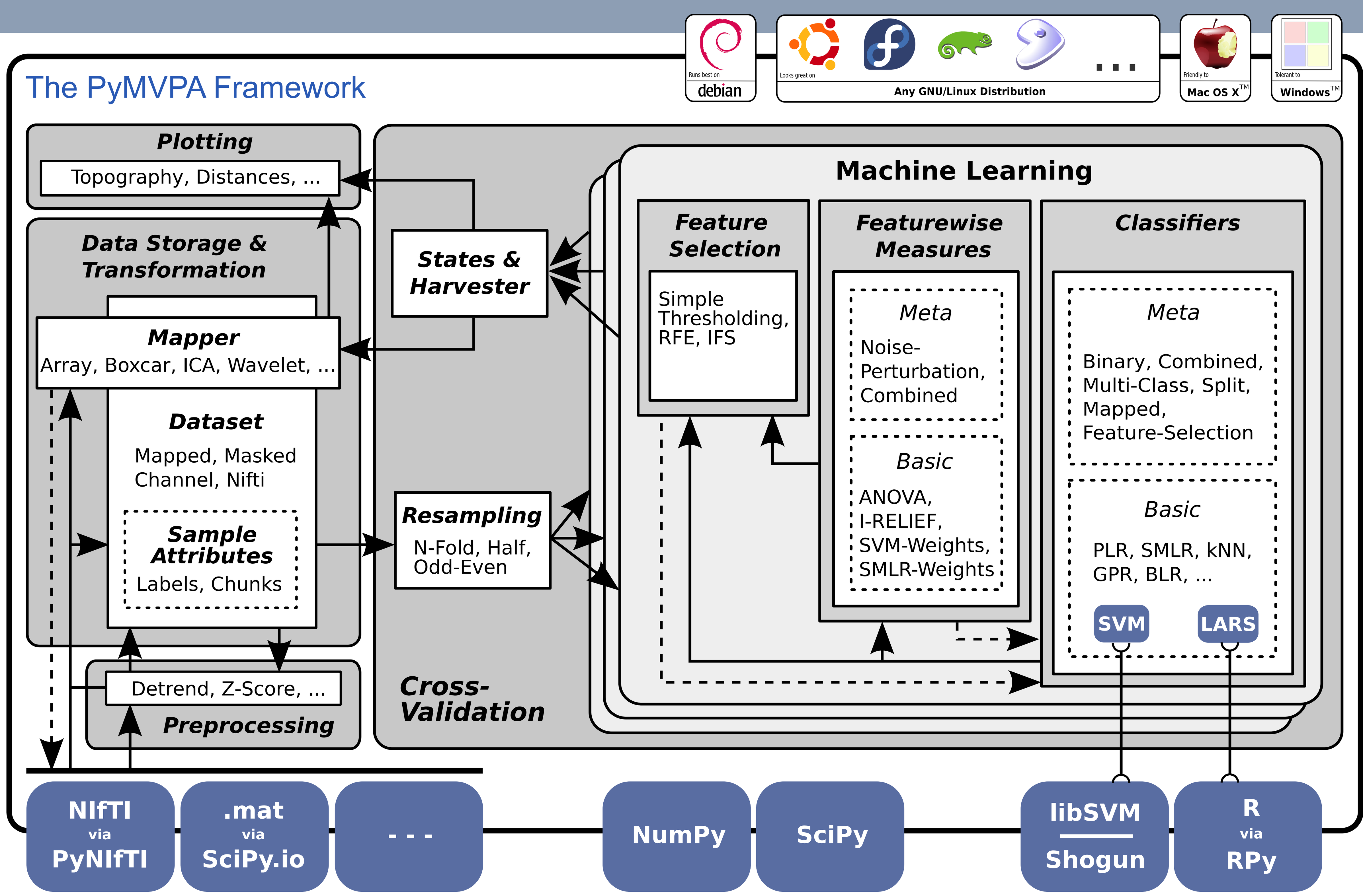
Why Use Machine-learning Based Analysis?

- Facilitates multivariate approach to discover distributed activation patterns
- Encourages model testing
- Provides direct quantifiable link between experimental conditions and fMRI data (O'Toole et al., 2007)
- Optimizes the analysis of high resolution imaging data (Kriegeskorte et al., 2007)
- Benefits from inter-disciplinary methodological developments

What Are The Cardinal Features of PyMVPA?

- Flexible framework to access existing machine-learning software
- Concise scripting interface for human-readable and verifyable code
- Modular architecture allowing extensions in multiple languages
- Portable code running everywhere from mainframes to cell phones
- Data modality independent, but not data modality ignorant
- **Free** and **open-source** software (FOSS)

The PyMVPA Framework



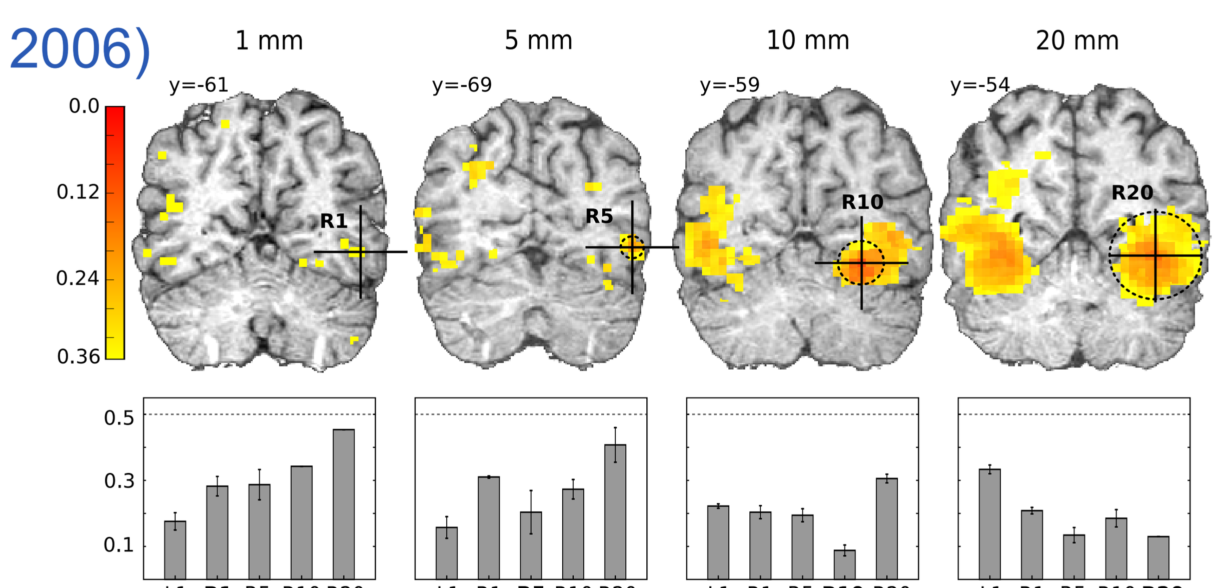
Example: Searchlight Analysis (Kriegeskorte et al., 2006)

```
attr = SampleAttributes('sample_attr_filename.txt')
dataset = NiftiDataset(
              samples='subj1_bold.nii.gz',
              labels=attr.labels,
              chunks=attr.chunks,
              mask='subj1_roi_mask.nii.gz')
```
LOAD DATA

```
cv = CrossValidatedTransferError(
         TransferError(LinearCSVMC()),
         OddEvenSplitter())
sl = Searchlight(cv, radius=5)
```
SET UP ANALYSIS PIPELINE

```
sl_map = sl(dataset)
```
RUN IT

```
dataset.map2Nifti(
    array(sl_map)).save('slight_5mm.nii.gz')
```
STORE RESULTS



CATS vs. SCISSORS classification error for four searchlight radii (Hanke et al., 2008), using a dataset of one participant from Haxby et al. (2001).

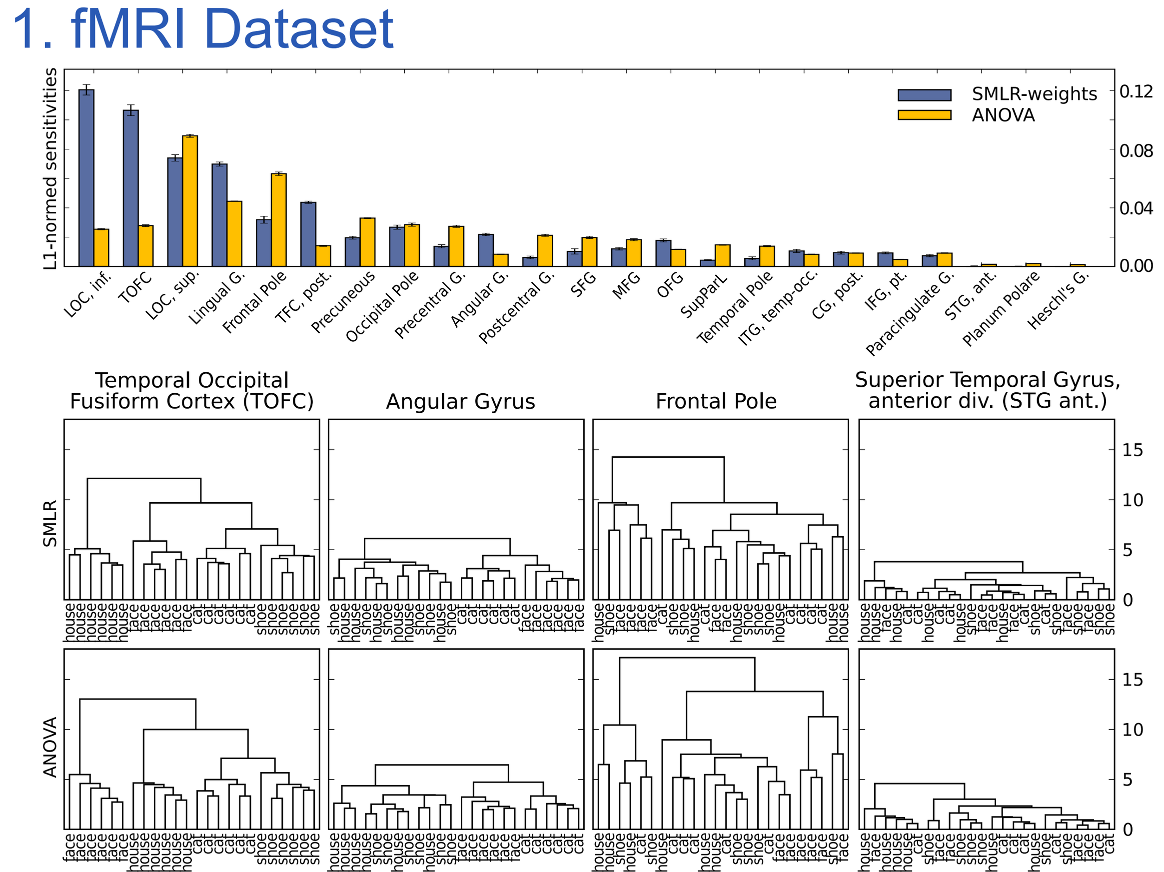
Uniform Analysis of Three Different Neuroimaging Data Modalities With *Identical* Source Code

Sketched out below is a modality-independent cross-validation analysis in PyMVPA (the pre-processing steps are not shown). Various linear classifiers and two featurewise measures are used to compute sensitvity scores. The figures below demonstrate how these sensitivities can be put back into their orginal context for visualization and interpretation.

```
# some classifiers to test
clfs = {'SMLR': SMLR(lm=0.1),
        'lCSVM': LinearCSVMC(),
        'lGPR': GPR(kernel=KernelLinear())}
# define some pure sensitivities (or related measures)
sensanas={'ANOVA': OneWayAnova(),
          'I-RELIEF': IterativeReliefOnline()}
# to store all computed sensitivities
senses = []

# compute sensitivities
for k, v in sensanas.iteritems():
    sa = SplitFeaturewiseMeasure(v, NFoldSplitter(),
                                 enable_states=['maps'])
    # compute and store sensitivities for all splits
    sa.append((k, sa(dataset).maps, None, None))
```

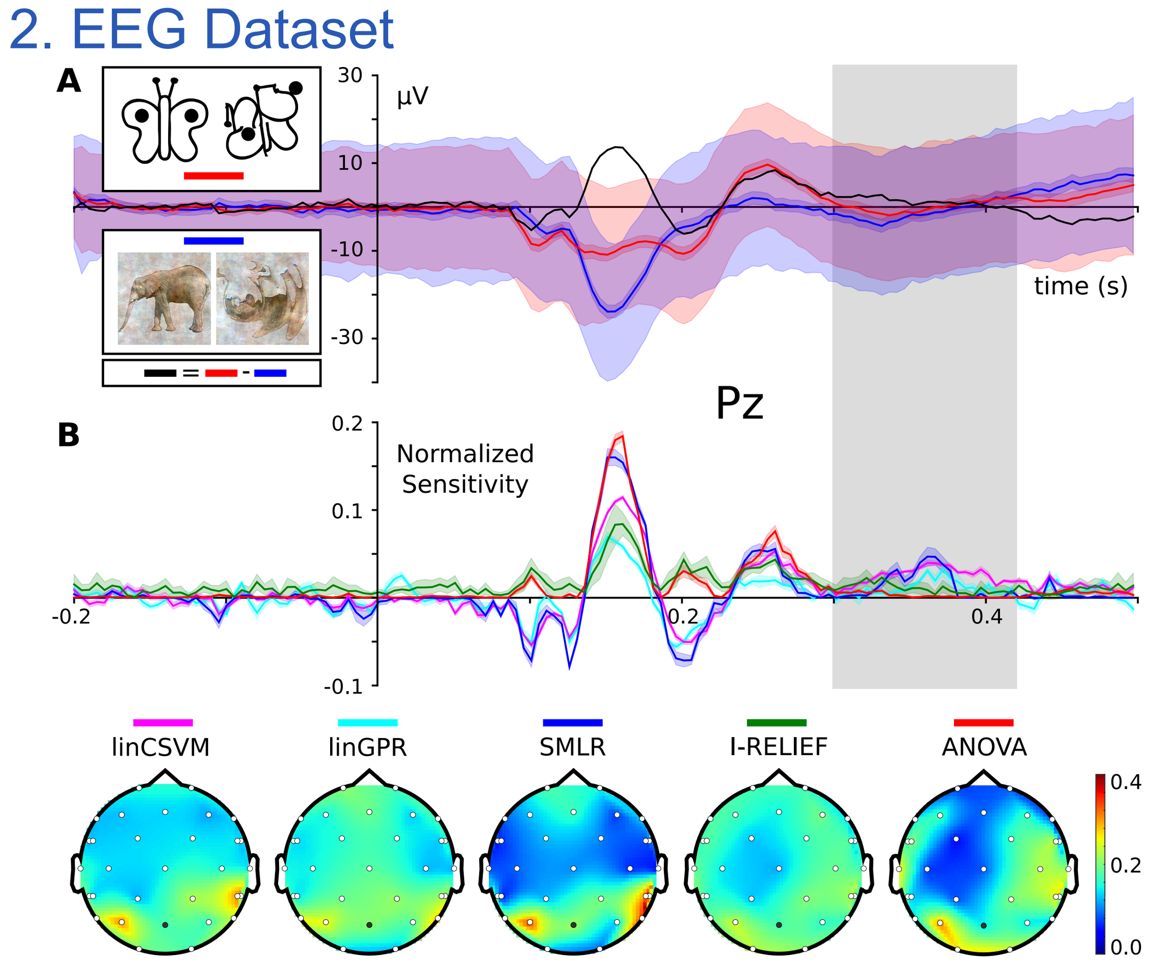
```
# run classifiers in cross-validation
for label, clf in clfs.iteritems():
    sclf = SplitClassifier(clf, NFoldSplitter(),
             enable_states=['confusion', 'training_confusion'])

    # Compute sensitivity, which in turn trains the sclf
    sensitivities = sclf.getSensitivityAnalyzer(
        # do not combine sensitivities across splits,
        # nor across classes
        combiner=None, slave_combiner=None)(dataset)

    # and store
    senses.append((label, sensitivities, sclf.confusion,
                   sclf.training_confusion))
```

1. fMRI Dataset



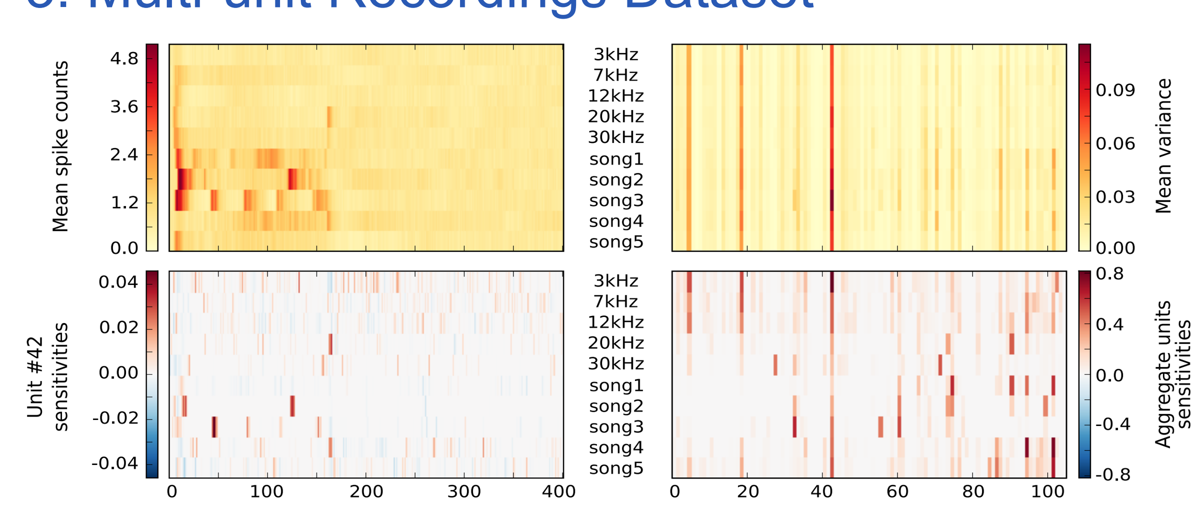
Sensitivity analysis of the Haxby et al. (2001) fMRI dataset (four-category subset). The upper part shows ROI-wise scores computed from SMLR classifier weights and ANOVA F-scores.The lower part shows dendrograms with clusters of block-averaged category samples (computed using squared Euclidean distances) for voxels with non-zero SMLR-weights and a matching number of voxels with the highest F-scores in each ROI.

2. EEG Dataset



Sensitivities for the classification of color and line-art stimulation conditions (EEG dataset from Fründ et al., 2008). **(A)** Event-related potentials (ERP). The shaded areas show the standard deviation and the 95% confidence interval around the mean ERP (black curve is the mean difference wave). **(B)** Normalized sensitivity measures (unit length vector norm over all channels). The head topography plots show the channel-wise sum over time of the absolute scaled sensitivities. The upper panel shows the scaled sensitivities plotted over time for the *Pz* electrode (dark dot on the head topographies). Note the similarity between the sensitivities and the ERP difference wave. Interestingly, for a time window around 350 ms after stimulus onset, all multivariate sensitivity measures assign a certain amount of weight, whereas the univariate ANOVA is flat at zero.

3. Multi-unit Recordings Dataset



Statistics of multi-unit extracellular recordings dataset (dataset from Luczak and Harris, unpublished study) and corresponding classifier sensitivities. The upper part shows basic statistics for each stimulus condition. The lower part shows the temporal sensitivity profile of a representative unit, and associated aggregate sensitivities of all units and stimulation conditions.

Upcoming Features And Future Development

- Convenient analysis of event-related datasets
- Cluster computing features, solely using FOSS (no licensing costs)
- Extended model selection capabilities for automatic tuning of classifier hyperparameters
- Graphical User Interface (GUI)

References

Fründ, I., Busch, N. A., Schadow, J., Gruber, T., Körner, U. & Herrmann, C. (2008). Time pressure modulates electrophysiological correlates of early visual processing. PLoS One, 3, e1675.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V. & Pollmann, S. (2008). PyMVPA: A Python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics, accepted.

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J. & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science, 293, 2425–2430.

Kriegeskorte, N., Goebel, R. & Bandettini, P. (2006). Information-based functional brain mapping. PNAS, 103, 3863–3868.

Kriegeskorte, N. & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage, 38, 649–662.

O'Toole, A. J., Jiang, F., Abdi, H., Penard, N., Dunlop, J. P. & Parent, M. A. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. Journal of Cognitive Neuroscience, 19, 1735–1752.