A very short introduction to multivariate pattern analysis (MVPA) for neuroscience

Michael Hanke & Yaroslav Halchenko

University of Magdeburg, Germany
Dartmouth College, USA

Giessen 2014
SPM via GLM (“Encoding model”)

Research Question → Experiment Design → Stimuli → Hypothesis Testing
SPM via GLM ("Encoding model")

$p(\text{brain activity}|\text{behavior})$

Research Question
Experiment Design
Stimuli

Hypothesis Testing

GLM

MVPA Intro

Giessen 2014
That’s not enough

Eric Kandel in *Principles of Neuroscience*

“The task of neural science is to explain behavior in terms of the activities of the brain.”

\[ p(\text{brain activity}|\text{behavior}) \neq p(\text{behavior}|\text{brain activity}) \]
Approach: Meta analysis and Bayes’ theorem

NeuroSynth is a platform for large-scale, automated synthesis of functional magnetic resonance imaging (fMRI) data extracted from published articles.

Our goal is to turn this:

Into this:

Yarkoni et al., Nature Methods, 2011; http://neurosynth.org
Why multivariate methods?

Haufe et al., 2014, NeuroImage

$H_2$ (Dartmouth; Magdeburg)  MVPA Intro

Giessen 2014  5 / 16
Why multivariate methods?
Why multivariate methods?
Pioneering work: visual objects

Haxby et al., Science, 2001

$H_2$ (Dartmouth; Magdeburg)  
MVPA Intro

Giessen 2014  7 / 16
k-Nearest Neighbours
k-Nearest Neighbours
k-Nearest Neighbours
MVPA approach: *Reverse the flow!*

\[ p(\text{behavior} \mid \text{brain activity}) \]
Which classifier?

classification

- SVC
- Ensemble Classifiers
- Naive Bayes
- Text Data
- Linear SVC
- Kernel approximation
- SGD Classifier

clustering

- Spectral Clustering
- GMM
- KMeans
- MiniBatch KMeans
- MeanShift
- VBGMM

scikit-learn algorithm cheat-sheet

regression

- SGD Regressor
- Lasso
- ElasticNet
- SVR(kernel='rbf')
- Ensemble Regressors
- RidgeRegression
- SVR(kernel='linear')

dimensionality reduction

- Randomized PCA
- Isomap
- Spectral Embedding
- LLE

START
Decision models – linear problem

10-Nearest-Neighbour

3-Nearest-Neighbour

GNB

GNB(common_variance=False)

LDA

Linear SVM

Logistic Regression

QDA

RBF SVM

Ridge Regression

SMLR

MVPA Intro

H₂ (Dartmouth; Magdeburg)
Decision models – non-linear problem

10-Nearest-Neighbour

3-Nearest-Neighbour

GNB

GNB(common_variance=False)

LDA

Linear SVM

Logistic Regression

QDA

RBF SVM

Ridge Regression

SMLR

H₂ (Dartmouth; Magdeburg)
Model appropriateness

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer. (free PDF copy)
Data representation – classification

Experimental Runs

Chunks

Experimental Conditions

Labels

fMRI Volumes

Samples

Time

Experimental Runs

Chunks

Experimental Conditions

Labels

fMRI Volumes

Samples

Time

$H_2$ (Dartmouth; Magdeburg)
Enough said... (for now)

Let’s see how we can do this...