Multivariate Pattern Analysis of fMRI
fMRI多變數模式分析發展與應用

盧家鋒 助理教授

國立陽明大學生物醫學影像暨放射科學系
alvin4016@ym.edu.tw

2019.1.8

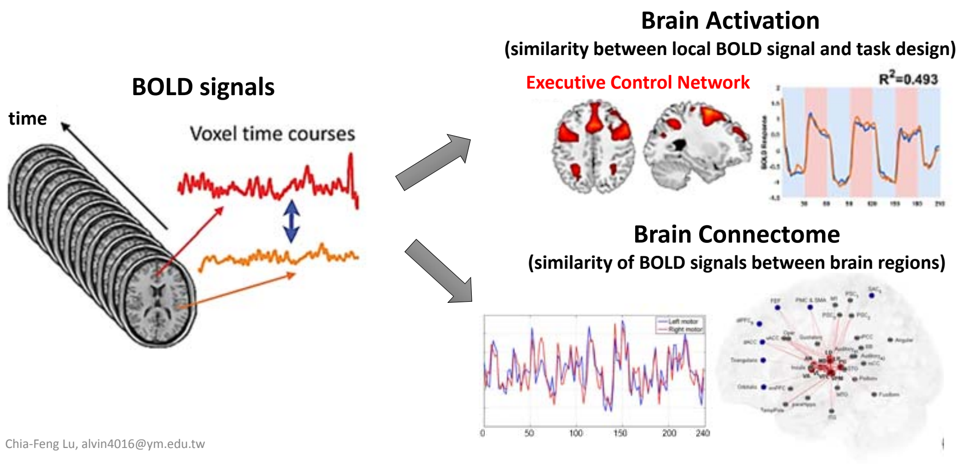
Schedule

- **13:00~13:50** Introduction of MVPA
- **14:00~14:50** Machine Learning in MVPA
- **15:00~15:50** Applications of MVPA
- **16:00~16:50** Perspectives on MVPA

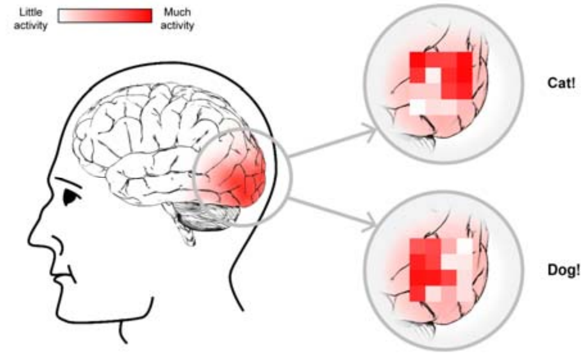
13:00~13:50 Introduction of MVPA

Lecture 1

From Brain Activation to Connectome



Decoding Activity Pattern of Brain



Looking at the pattern of activation within a brain area allows us to answer what the person is seeing.

Illustration by Pim Mostert http://blog.donders.ru.nl/?p=4361&lang=en

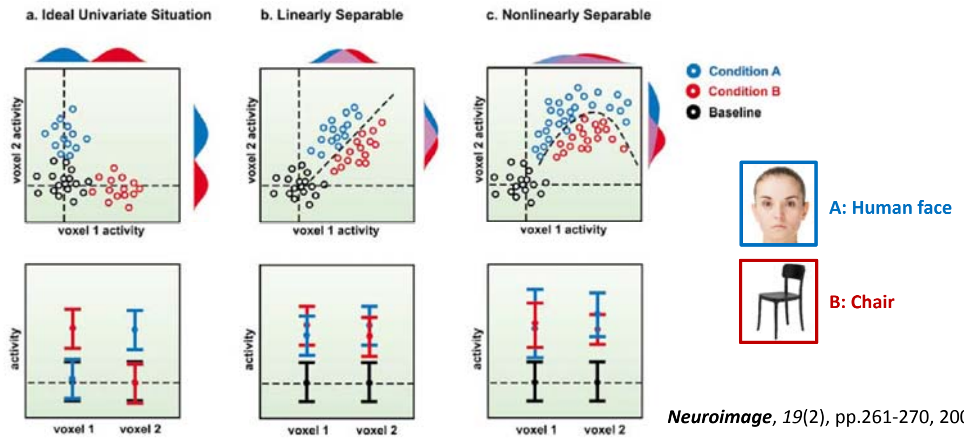
Background Articles of MVPA

- **[Review] James V. Haxby**, Multivariate pattern analysis of fMRI: The early beginnings. ***Neuroimage***, 2012.
- Haxby, J.V. et al., Distributed and overlapping representations of faces and objects in ventral temporal cortex. ***Science***, 2001.
- Cox, D.D. and Savoy, R.L., Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. ***Neuroimage***, 2003
- Haxby, J.V. et al., A common, high-dimensional model of the representational space in human ventral temporal cortex. ***Neuron***, 2011.

Brain Activation → Brain Decoding

- **Mass-univariate model-based analysis**
  - Analyze every voxel (~50,000) one at a time
  - General Linear Model, GLM (since 1995)
- **Multivoxel Pattern Analysis, MVPA**
  Multivariate Pattern Analysis, MVPA
  - Original version: correlation analysis
  - Machine learning: Support Vector Machine, SVM

Why we need multivariate analysis?



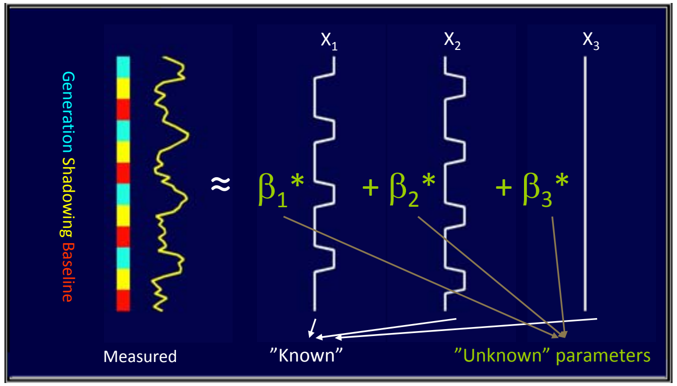
Neuroimage, *19*(2), pp.261-270, 2003.

Methodological Article of MVPA

- Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D. and Brovelli, A., 2012. Multivoxel pattern analysis for FMRI data: a review. ***Computational and mathematical methods in medicine***, *2012*.
- Lewis-Peacock, J.A. and Norman, K.A., 2013. Multi-voxel pattern analysis of fMRI data. ***The cognitive neurosciences***, pp.911-920.

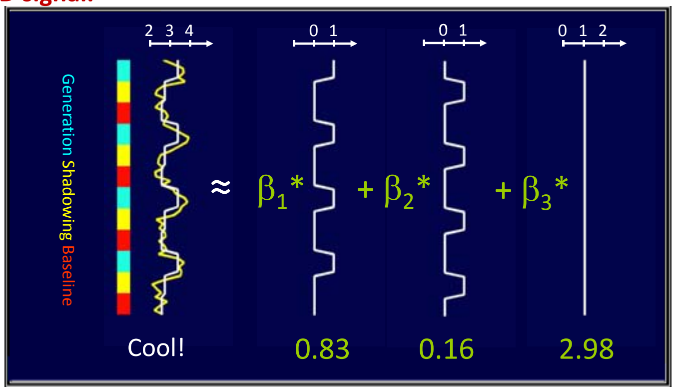
The Model of GLM

Finding the linear combination of these hypothetical time series "best" fits the data.



Parameter Estimation

Beta value represents the association between a condition design and the measured BOLD signal.

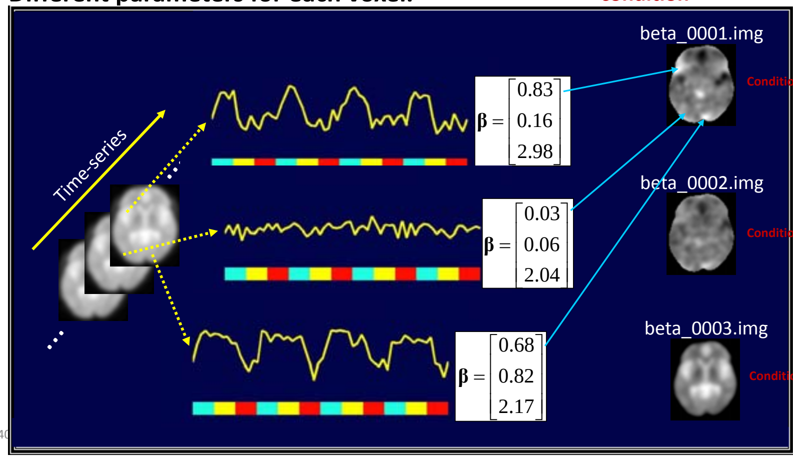


Parameter Estimation

Same model for all voxels.

Different parameters for each voxel.

Beta map for each task condition



Major limitations of GLM

- The basic assumption that the covariance across neighboring voxels is not informative about the cognitive function under examination.
- Such covariance is considered as uncorrelated noise and normally reduced using spatial filters that smooth BOLD signals across neighboring voxels.
- Additionally, the GLM approach is inevitably limited by the model used for statistical inference.

Major limitations of GLM

- Ignores correlation between voxels
- Cannot capitalize on pooling information over voxels
- Fails to capture “distributed” neural codes

- Multiple comparisons Problem
- Model-based inference

Answers to the Questions

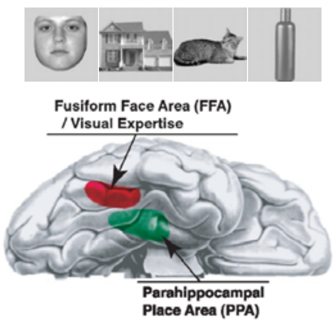
Brain activation to the corresponding task.
- Where is the motion area?
- Where is the face recognition area?

Specialized areas
⇕
Encoding patterns

Brain encoding to the representation of stimulus classes.
- What are the varying brain states in an area?
- How do brain cortices encode different types of information?

Origin of MVPA

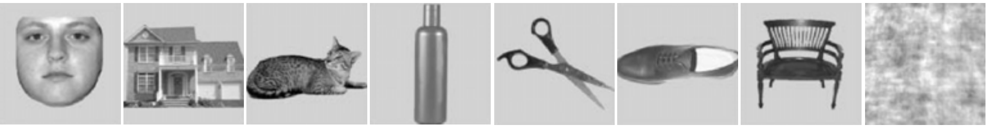
Three hypothesis made by James V. Haxby, 2001.
- Each object category would evoke a distinct pattern of response in ventral temporal cortex.
- These distinctive patterns would not be restricted to category-selective regions, such as the FFA (face) and PPA (other objects).
- Neural activity patterns within category-selective regions would carry information that discriminates between categories.



Origin of MVPA

Object form topography

- The ventral temporal cortex has a topographically organized representation of attributes of form that underlie face and object recognition.
  - Can produce unique representations for a virtually unlimited number of categories.



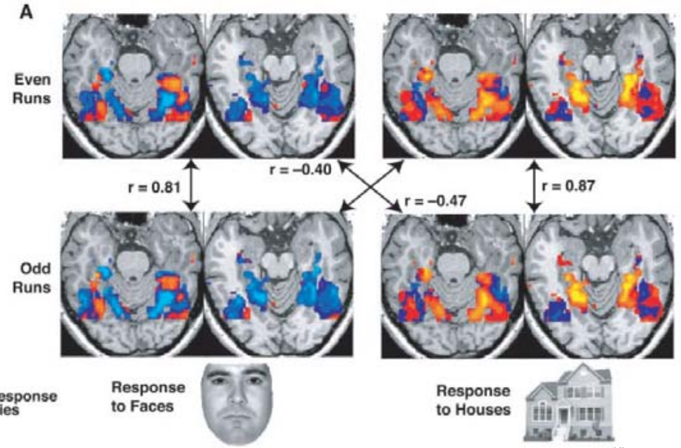
Science, *293*(5539), 2425-2430, 2001.

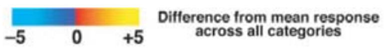
Origin of MVPA

Science, *293*(5539), 2425-2430, 2001.

Split-sample cross-correlation (odd vs. even runs)

- The response to that category should be more similar to each other than to responses to different categories.





Model-free MVPA

- Investigating the functional role of distributed patterns of neural activity without assuming a specific model.
- MVPA involves searching for highly reproducible spatial patterns of activity that differentiate across experimental conditions.
- Machine learning and pattern recognition algorithms.
  - a classifier attempts to capture the relationships between spatial patterns of fMRI activity and experimental conditions.

Toolbox of MVPA

- Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollman, S., 2009. PyMVPA: a Python toolbox for multivariate pattern analysis of fMRI data. ***Neuroinformatics*** 7, 37–53.
- An MVPA toolbox using Matlab (the Princeton MVPA toolbox) (http://code.google.com/p/princeton-mvpa-toolbox/).
- Oosterhof, N.N., Connolly, A.C. and Haxby, J.V., 2016. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. ***Frontiers in neuroinformatics***, *10*, p.27.

14:00~14:50 Machine Learning in MVPA

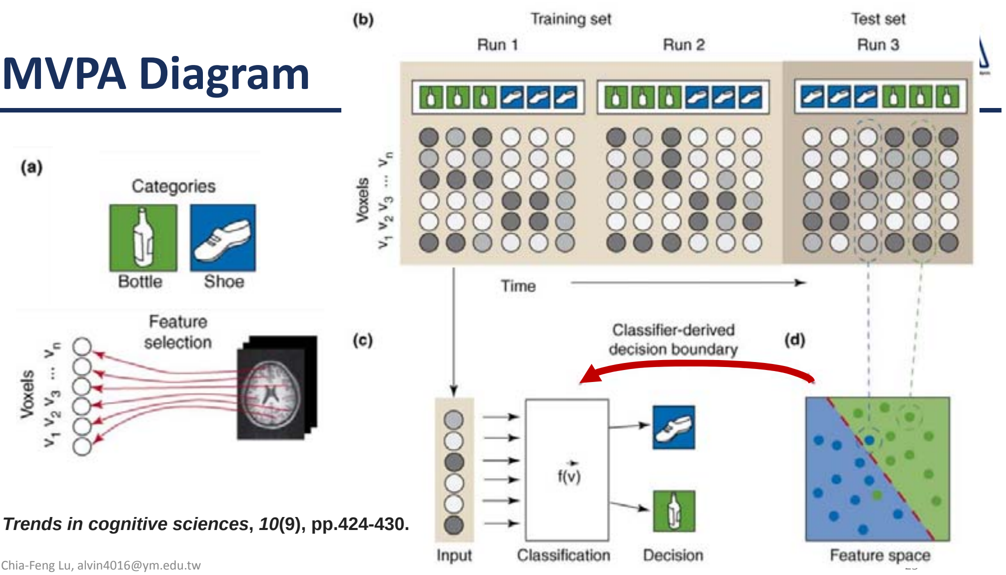
Lecture 2

MVPA: A Classification Problem

- Classification consists in determining a decision function f that takes the values of various “features” in a data “example” x and predicts the class of that “example.”

- An “**example**” may represent a given trial in the experimental run.
- The “**features**” may represent the corresponding fMRI signals in a cluster of voxels.
- The experimental conditions may represent the different “**classes**”.

MVPA Diagram



***Trends in cognitive sciences*, *10*(9), pp.424-430.**

Learning Process – Identify Key Features

The key features of a CAR?

(Compared to a motorcycle)

Learning Process – Identify Key Features

Mushrooms or buns?

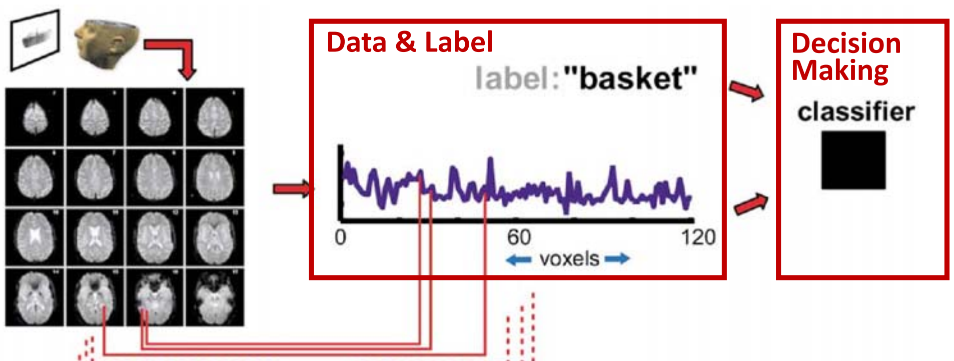
Weights between factors

- Appearance 20%
- Smell 5%
- **Taste 75%**

Supervised Learning/Classification

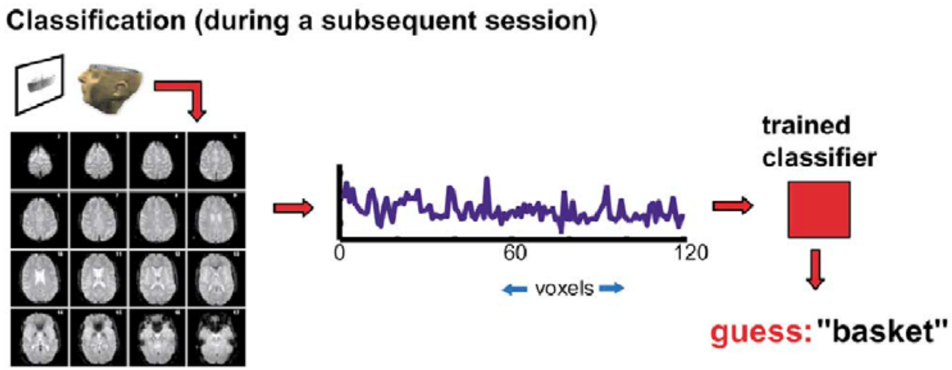
- To obtain the decision function *f* , data (i.e., examples and the corresponding class labels) must be split into two sets: "**training set**" and "**test set**."
- Training consists of modeling the relationship between the features and the class label by assigning **a weight *w*** to each feature.

Training phase (Teaching)



Neuroimage, *19*(2), pp.261-270, 2003.

Testing phase



Neuroimage, *19*(2), pp.261-270, 2003.

Classifiers

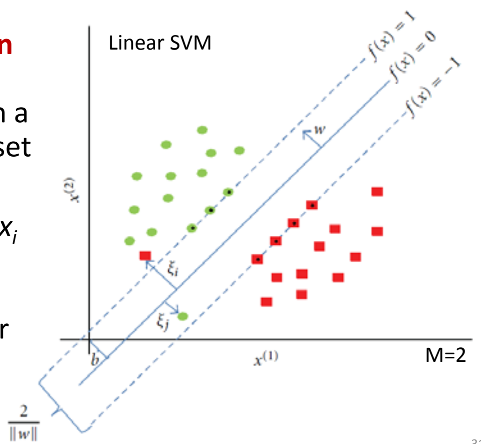
- Linear discriminant analysis
- **Support vector machines**
- Logistic regression
- K nearest neighbors
- Decision trees
- Neural networks

Support Vector Machine (SVM)

Alice Zhao, https://youtu.be/N1vOgolbjSc

Support Vector Machine (SVM)

- The goal is to estimate **a decision boundary (a hyperplane)** that separates with maximum margin a set of positive examples from a set of negative examples.
- Each example is an input vector x_i ($i = 1, . . . ,N$) having M features (i.e., x_i in R^M) and is associated with one of two classes $y_i = -1$ or $+1$.



Variable Description

In fMRI research,…

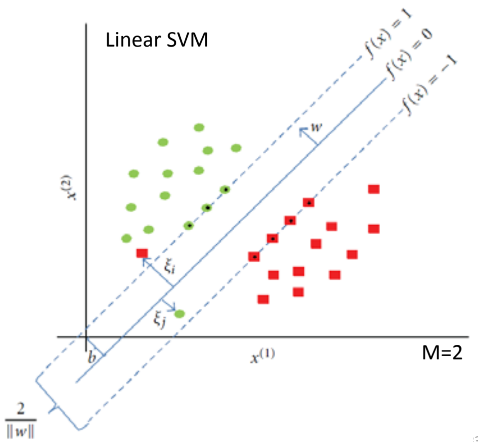
- the data vectors x_i contain BOLD values at discrete time points (or averages of time points) during the experiment;
- features could be a set of voxels extracted in each time point;
- $y = -1$ indicates condition A, and $y = +1$ indicates condition B.

Discriminant Function

- The SVM produces the discriminant function *f* with the largest possible margin:

$$f(x) = w \cdot x + b$$

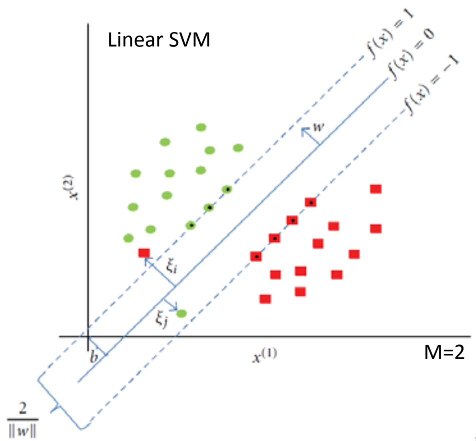
- *w* is the normal weight vector of the separating hyperplane;
- *b* is referred to as the "bias."



Discriminant Function

- SVM attempts to find the optimal hyperplane $w \cdot x + b = 0$ which maximizes the margin magnitude, that is, it finds *w* and *b* by solving the following *primal* optimization problem:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

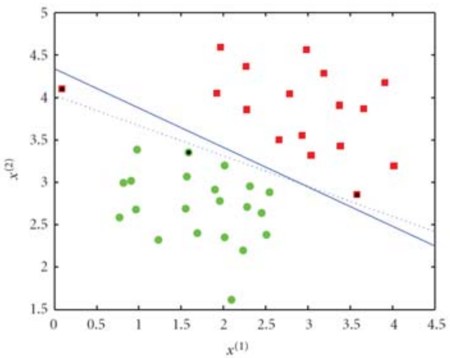


Parameters of SVM

- To control the trade-off between the hyperplane complexity and training errors, a penalty factor *C* is introduced.
- The *primal* optimization problem becomes

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$$

Blue line: C=0.1, dashed line: C= 10000.



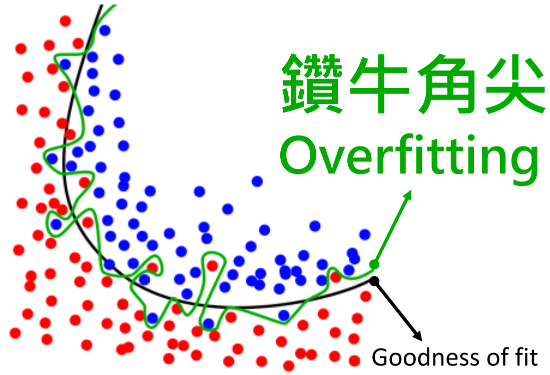
Acceptable classifier performance ⇔ overfitting

Issue of Overfitting

- Model Performance ⇔

  Model Generalization

- Very large *C* does not allow any training error ➔ overfitting.



Nonlinear SVM – kernel trick

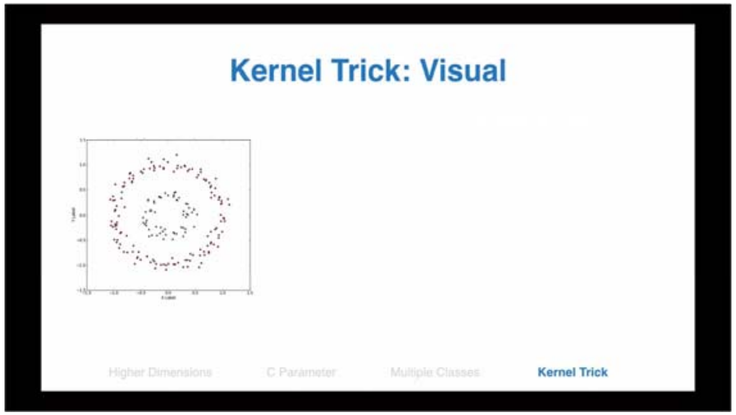
- Nonlinear SVMs are often used for discrimination problems when the data are nonlinearly separable.
- Vectors are mapped to a high-dimensional feature space using a function $g(x)$.

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i g(x_i) \cdot g(x) + b$$

- It allows a nonlinear operator to be written as a linear one in a space of higher dimension.

Nonlinear SVM – kernel trick

Alice Zhao, https://youtu.be/N1vOgolbjSc



Pros & Cons of SVM

- **Pros**
  - Good at dealing with high dimensional data (multiple voxels)
  - Works well on small data sets (a small number of trials)
- **Cons**
  - Picking the right kernel and parameters can be computationally intensive.

Components of Machine Learning

MVPA of fMRI

Feature Selection & Reduction → Classifier → Performance Evaluation

Feature Selection & Reduction
- RFE algorithm
- Most activated voxels
- Principal component analysis (PCA)

Classifier
- Linear discriminator
- Support vector machines
- Neural networks

Performance Evaluation
- ROC Curve
- Cross validation

Feature selection

- Variables that contain little information about the discrimination being made only add unrelated noise to the classifier and degrade performance.
- Classifier performance may be improved by reducing the data dimensionality (voxel numbers) or by selecting a set of discriminative features.

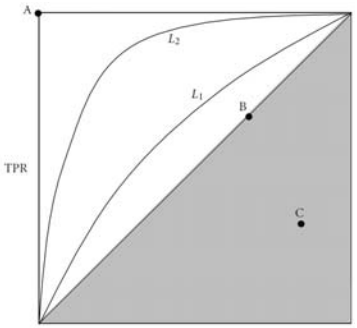
Recursive features elimination (RFE)

- For each voxel selection level, the RFE consists of two steps.
- First, an SVM classifier is trained on a subset of training data using the current set of voxels.
- Second, a set of voxels is discarded according to their discriminative weights as estimated during training.
- Data used as test are classified, and generalization performance is assessed at each iteration.

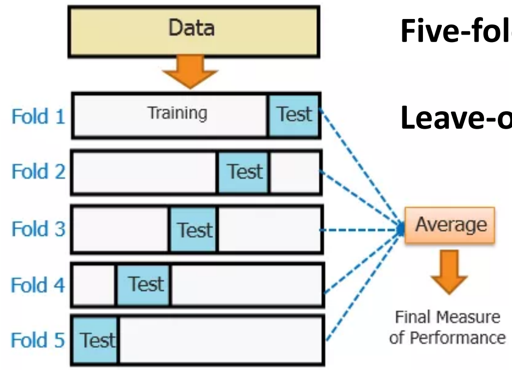
Model Performance: ROC Curve

- Receiver Operating Characteristic (ROC) Curve
- The ROC curve is formed by plotting true positive rate (TPR) over false positive rate (FPR) defined both from the *confusion* matrix by

		p	n
Predicted class	Y	TP (True positives)	FP (False positives)
	N	FN (False negatives)	TN (True Negatives)
		n^+	n^-



Model Performance & Selection



Five-fold cross-validation

Leave-one-out cross-validation

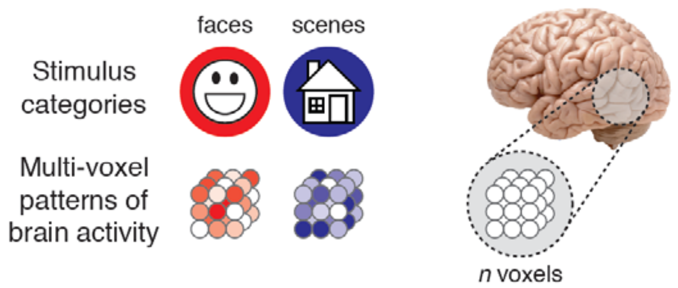
https://blog.contactsunny.com/data-science/different-types-of-validations-in-machine-learning-cross-validation

15:00~15:50 Applications of MVPA

Lecture 3

Processing Steps of MVPA

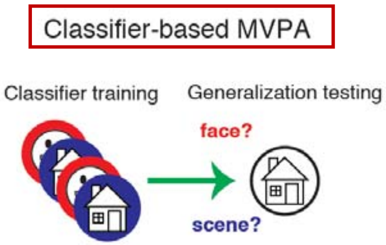
Step 1: Feature selection and pattern assembly



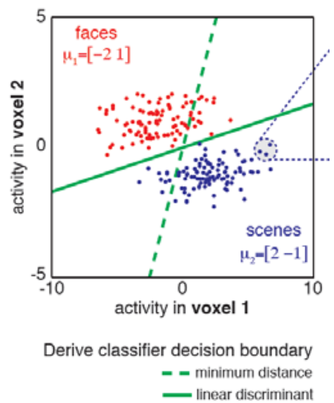
The cognitive neurosciences, pp.911-920, 2013.

Processing Steps of MVPA

Step 2: Multi-voxel pattern analysis (MVPA)



The cognitive neurosciences, pp.911-920, 2013.



Experimental Design

Block design

- Subjects were shown images belonging to each of ten categories organized into ten 20s blocks, with 20-s fixation blocks at the beginning (1st block), middle (7th block), and end (13th block) of each run. During each block, 10 stimuli were shown for 2 s each.

Event-related design

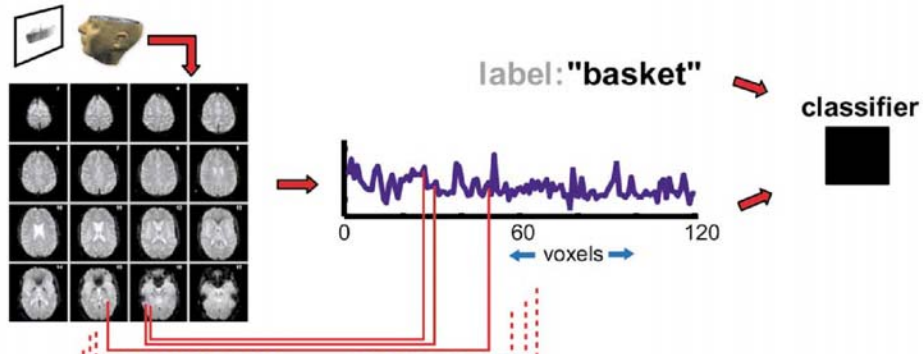
- We employed a two-condition slow event-related experiment. Events last 500 ms, and their onsets are separated by 16 s. There are 20 events per condition. The entire simulated experiment thus lasts 10 min 40 s. The condition sequence is random. The temporal resolution is one volume per 2 s.

Preprocessing steps for MVPA

- Slice timing and realignment (motion-correction) are demanded for essential imaging correction.
- No explicit temporal or spatial smoothing should be performed.
- Spatial transformation into a standard space prior to analysis is optional.

- Individual runs should be normalized by subtracting the mean activity during the fixation blocks.
- [optional] Data from individual blocks were averaged together over a window starting two acquisitions after the beginning of the block and ending at the end of the block.

Processing flow



Neuroimage, *19*(2), pp.261-270, 2003.

Univariate-based ROIs of MVPA

- Cox, D.D. and Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. ***Neuroimage***, *19*(2), pp.261-270.

- **A pattern recognition problem is at its root a "brain-reading" problem.**
  - Given a pattern of brain activity across space, a pattern recognition approach seeks to infer what percept a subject was experiencing.

Univariate-based ROIs of MVPA

One-way ANOVA

- Identify voxels that vary significantly across at least one of the categories of stimuli ($P < 0.05$, Bonferroni corrected for multiple comparisons).

Correlation analysis

- Identify voxels that were significantly correlated with the predictor (boxcar convolved with HRF, $r > 0.4$) were deemed "object processing areas."

Searchlight Approach

- Kriegeskorte, N., Goebel, R. and Bandettini, P., 2006. Information-based functional brain mapping. ***Proceedings of the National Academy of Sciences***, *103*(10), pp.3863-3868.

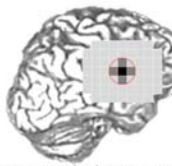
- **Where in the brain does the activity pattern contain information about the experimental condition?**
  - Rather than asking where in the brain does the average activity changes across experimental condition.

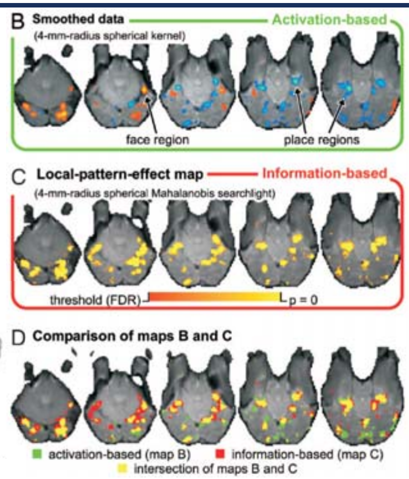
Searchlight Approach

When the **target** is to find…	**Assumption**	**Combination** of local signals
Activated Regions →	**Activate a region as a whole** →	**Local averaging** (spatial smoothing)
	V.S.	
Informative Regions →	**Activity patterns as distributed representations** →	**Multivariate statistics** (compare activity patterns among conditions)
	Salt-and-pepper fine structure	

Searchlight Approach

- Scan the brain with a **spherical multivariate "searchlight"** centered on each voxel in turn.
  - An optimal radius of 4mm contains 33 2-mm-isotropic voxels.
- The resulting map shows **how well the multivariate signal in the local spherical neighborhood differentiates the experimental conditions.**
  - Average absolute t value
  - Mahalanobis distance

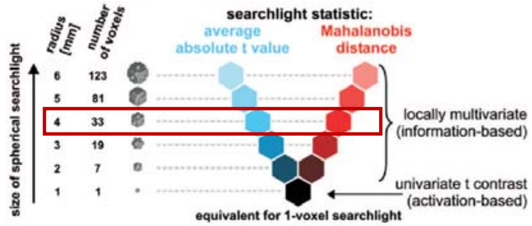


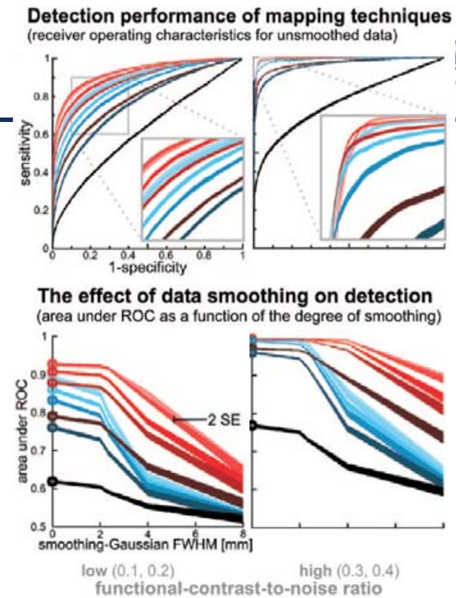


PNAS, 103(10), 3863-3868.

Searchlight Approach

The searchlights yielding optimal **with 4- or 5-mm radius.**



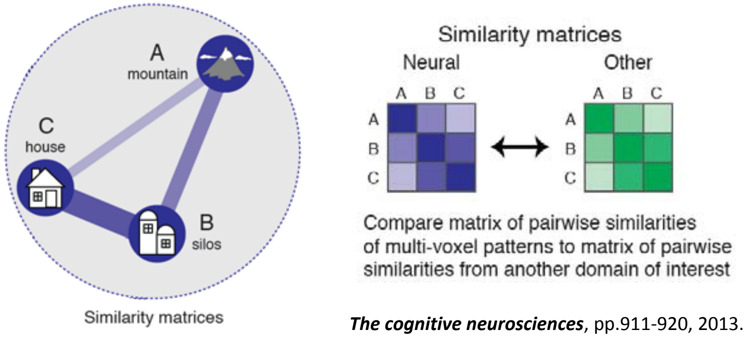


PNAS, 103(10), 3863-3868.

Essential Steps of MVPA

Step 2: Multi-voxel pattern analysis (MVPA)

Pattern-similarity MVPA



The cognitive neurosciences, pp.911-920, 2013.

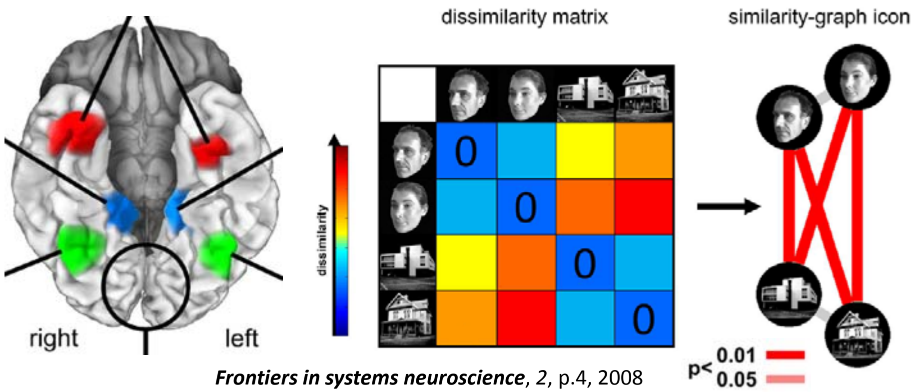
Pattern Similarity Analysis

- Kriegeskorte, N., Mur, M. and Bandettini, P.A., 2008. Representational similarity analysis-connecting the branches of systems neuroscience. ***Frontiers in systems neuroscience***, *2*, p.4.

- Make inferences about the similarity of mental concepts based on the similarity of patterns of brain activity.
- Regions that differentiate items within a category should show greater pattern similarity between two instances of the same item, compared to two distinct items from the same category.

Pattern Similarity Analysis

Compared between regions, species, mental states, and diseases.



Frontiers in systems neuroscience, *2*, p.4, 2008

16:00~16:50 Perspectives on MVPA

Lecture 4

MVPA: To what end?

- Can map representational code of stimulus classes (e.g. visual object categories).
- Can decode and predict "cognitive states" related to attention, memory, and perception.
- Can be used to detect awareness in patients with minimal consciousness.
- Can be used to classify disease states (Alzheimer's disease, concussion, etc.).
- Can be used to detect lies (murderers and non-murderers).

Quote from presentation of Bradley Buchsbaum, 2016.

Localization vs. Multivariate Sensitivity

- Univariate analysis has maximal resolution for estimating "activity."
- MVPA trades localization for sensitivity and inference power.

- For MVPA, however, must decide on a spatial scale.
  - 10 voxels, 1000 voxels, or whole brain?
  - Searchlight approach or univariate-based ROIs
- Depends on question and type of experiment.

Cautionary Notes for information mapping

- A brain region contains information about a particular cognitive distinction does not necessarily imply that this region is involved in guiding behavior based on that distinction (e.g., Williams, Dang, & Kanwisher, 2007).
- Information-mapping may produce false positive results; MVPA can be more susceptible than univariate analysis to experimental confounds (e.g., task difficulty) (Todd, Nystrom, and Cohen, 2013).

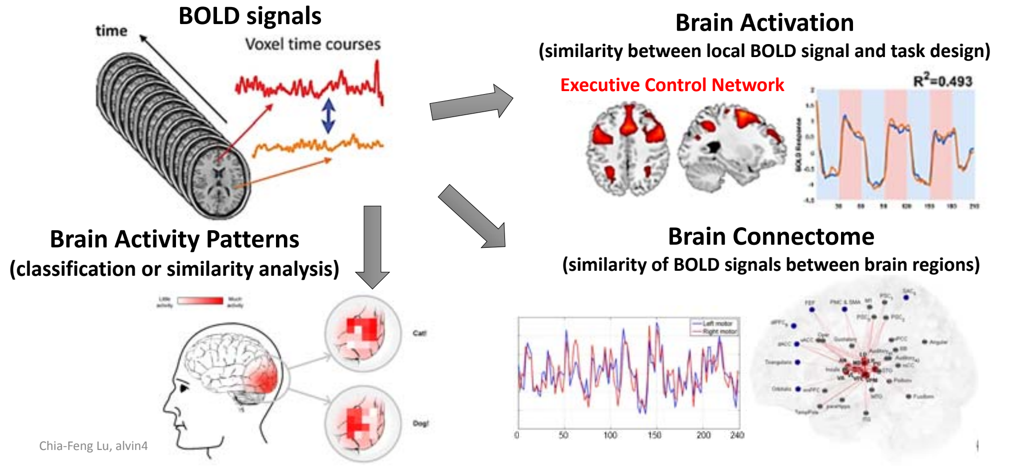
Cautionary Notes for information mapping

- Multivariate decoding is not necessarily more sensitive than univariate decoding (Jimura and Poldrack, 2012).
  - If the underlying signal has a coarse spatial scale, then univariate approaches using spatial smoothing at this scale will outperform MVPA.
- The extra parameters used to model the data in MVPA can lead to overfitting (Kriegeskorte et al., 2006).

Cautionary notes for pattern-similarity analysis

- Pattern similarity analyses do not compute weights for each voxel - these analyses treat all voxels as equally important.
  - more susceptible to contamination from noisy features than classifiers.
- Pattern similarity results can be influenced by univariate effects.
  - Voxels activate together for remembered items (even they are distinct objects) → increase of the average pattern similarity between remembered items.
  - One might interpret this effect in terms of neural representations “converging” in representational space between distinct objects.

From Brain Activation to Connectome



THE END

alvin4016@ym.edu.tw