Machine Learning in fMRI

Alexandre Savio, Maite Termenón, Manuel Graña[1]

[1]Computational Intelligence Group
University of the Basque Country

Outline

Experimental Design

- As a general principle the experimenter has to decide in as much detail as possible what he/she wants from the experiment.
- The scientific question may not be suitable for neuroimaging, and this very basic point must be addressed at the beginning of every research project.
- fMRI: This stage involves the formulation of a hypothesis and this necessarily will influence the scheme adopted for the cognitive task conditions (stimulus presentation strategies, resting state), and image acquisition parameters.

Data Acquisition

- This is a process where we don't take part but it determines our work.
- The way data are captured (noise, orientation), the format of the captured images (nifti, dicom), ...

Outline

Image Registration

- In image processing one is often interested not only in analyzing one image but in comparing or combining the information given by different images.
- The task of image registration is to find an optimal geometric transformation between corresponding image data using a template or a reference image

Outline

Image Segmentation

- Depending on the problem we are dealing with, we can divide the brain into different regions and discard all the information we are not interested in.
- For example, it is common to remove the skull because it does not give relevant information and can complicate the classification process.

Outline

Feature Extraction

- It is a process of transforming image data into feature vectors.

$$\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$$

$$\mathbf{x_i} = \left[x_i^{(1)}; x_i^{(2)}; \ldots; x_i^{(d)}\right]$$

where n is the number of samples and d is the number of components of each feature vector.

- Geometrically, every $\mathbf{x_i}$ represents a point in a d-dimensional feature space.

Outline

Curse of Dimensionality

- This phenomena occurs (even more in fMRI) when the dimensionality of feature space is too high compared with the number of samples $(d \gg n)$.
- This leads to a degradation on the performance of the analysis process.
- When this happens, it is convenient to reduce the feature space dimensionality keeping all the relevant information.
- There are two common ways of doing this:
  - Feature Selection.
  - Feature Transformation.

Feature Selection

- This technique selects a subset of the components of the feature space.
- There are multiple ways to proceed with the feature selection process.

Selection Criteria

- Pearson correlation.
- Average intensity in multiple TRs.

Feature Transformation

- It consists of projecting the data onto a new feature space of fewer dimensions through a functional mapping

$$f : \mathbb{R}^d \rightarrow \mathbb{R}^m$$

such that $m < d$.

Feature Transformation techniques

- PCA - Principal Component Analysis: A transformation where the data set receives a new coordinate system, in which new axes follow the direction of greatest variance in the data set.
- ICA - Independent Component Analysis: It is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals.
- LICA - Latice Independent Component Analysis: It follows the same philosophy as ICA but in the Lattice Algebra framework.

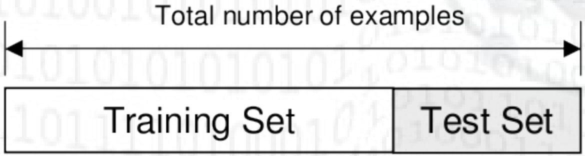
Outline

Resampling

- Hold-out
- Cross Validation:
    - k-Fold
    - Leave-one-out

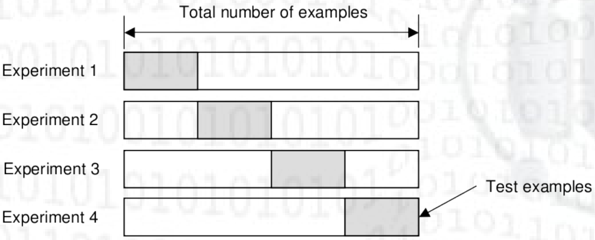
Hold-out

- Split dataset into two groups:
  - Training set: used to train the classifier.
  - Test set: used to estimate the error rate of the trained classifier



k-Fold Cross Validation

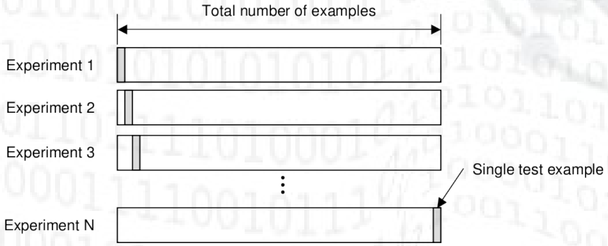
- For each of K experiments, use K-1 folds for training and the remaining one for testing.



- The advantage of K-Fold Cross-validation is that all the examples in the dataset are eventually used for both training and testing.
- Common choice for K-Fold Cross-validation is K=10.

Leave-one-out Cross Validation

- Leave-one-out is the degenerate case of K-Fold Cross Validation, where K is chosen as the total number of examples.
- For each experiment use N-1 examples for training and the remaining example for testing, then N experiments will be performed.



Outline

Classification

- Classification is the process of labelling each sample as one of the k available classes.
- There are two general types of classification:
  - Supervised classification.
  - Unsupervised classification.

Supervised Classification

- The process of using samples of known informational classes (training sets) to classify pixels of unknown identity.
- Most common algorithms:
    - LDA (Linear Discriminant Analysis), kNN (k-Nearest Neighbour).
    - ML (Maximum Likelihood), MAP (Maximum A Posteriori).
    - MLP (Multi Layer Perceptron), RBF (Radial Basis Function).
    - Non Linears: SVM (Support Vector Machine), RVM (Relevance Vector Machine).

Unsupervised Classification

- The identification of natural groups or structures/patterns with no information about the classes they belong to.
- Most common algorithms:
  - k-MEANS: It is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.
  - SOM: Self Organizing Maps. It is a type of artificial neural network that produce as low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map.

Unsupervised Classification

- Dendrograms: It is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering. Dendrograms are often used in computational biology to illustrate the clustering of genes or samples.
- LBG: Linde-Buzo-Gray Algorithm. It is a vector quantization algorithm to derive a good codebook. It is similar to the k-means method in data clustering.

Outline

Contingency Matrix

		actual value *p*	actual value *n*	total
prediction outcome	***p'***	True Positive	False Positive	P'
prediction outcome	***n'***	False Negative	True Negative	N'
	total	P	N	

Measures of fit

- There are a big variety of measures to give as validation. The most used in the medical literature are Accuracy, Sensitivity and Specificity.
- $TP + FP + FN + TN = N$
- Accuracy:
  - $Acc = (TP + TN)/N$
- True Positive Rate or Sensitivity or Recall:
  - $TPR = \frac{TP}{TP+FN} = Sensitivity$
- False Positive Rate:
  - $FPR = \frac{FP}{FP+TN} = 1 - Specificity$

Conclusions

- The goal of the analysis is to validate the work hypothesis.
- It is, to conclude if the studied phenomena (for instance, if the patient has a neurodegenerative disease or not) could be discerned from a fMRI data image using classification techniques.

Acknowledgements I

- Validation Lecture - Ricardo Gutierrez-Osuna - Intelligent Sensor Systems - Wright State University (http://courses.cs.tamu.edu/rgutier/ceg499_s02/l13.pdf)

References I

- Tools:
  - Quick-R: http://www.statmethods.net/advstats
  - PyMVPA: http://www.pymvpa.org
  - LibSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm/
  - SVM-light: http://svmlight.joachims.org/
  - Weka: http://www.cs.waikato.ac.nz/ml/weka/

- Pechenizkiy, M; Tsymbal, A; Puuronen, S: "PCA-based Feature Transformation for Classification: Issues in Medical Diagnostics". Proceedings. 17th IEEE Symposium on Computer-Based Medical Systems, 2004. CBMS 2004. 535 - 540.