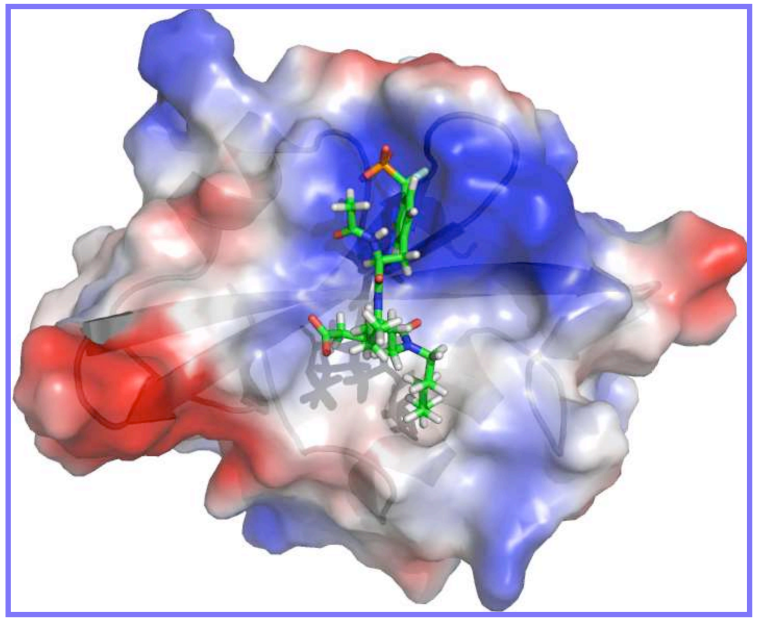
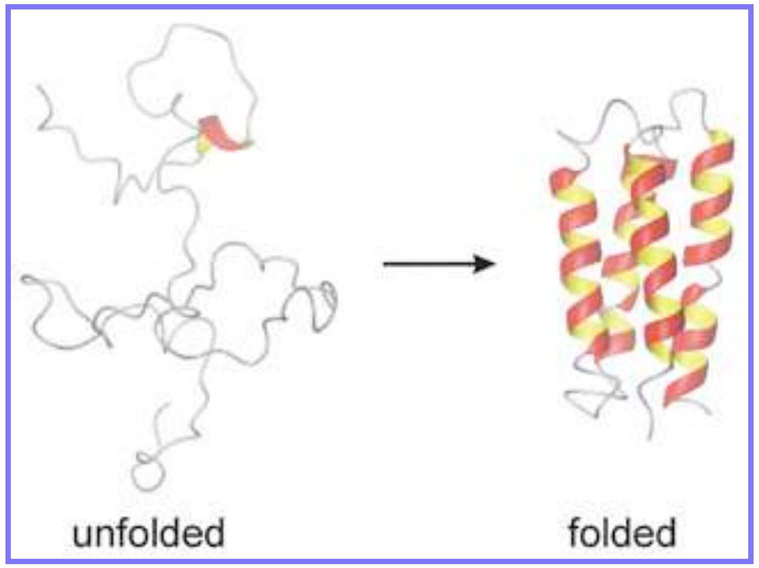
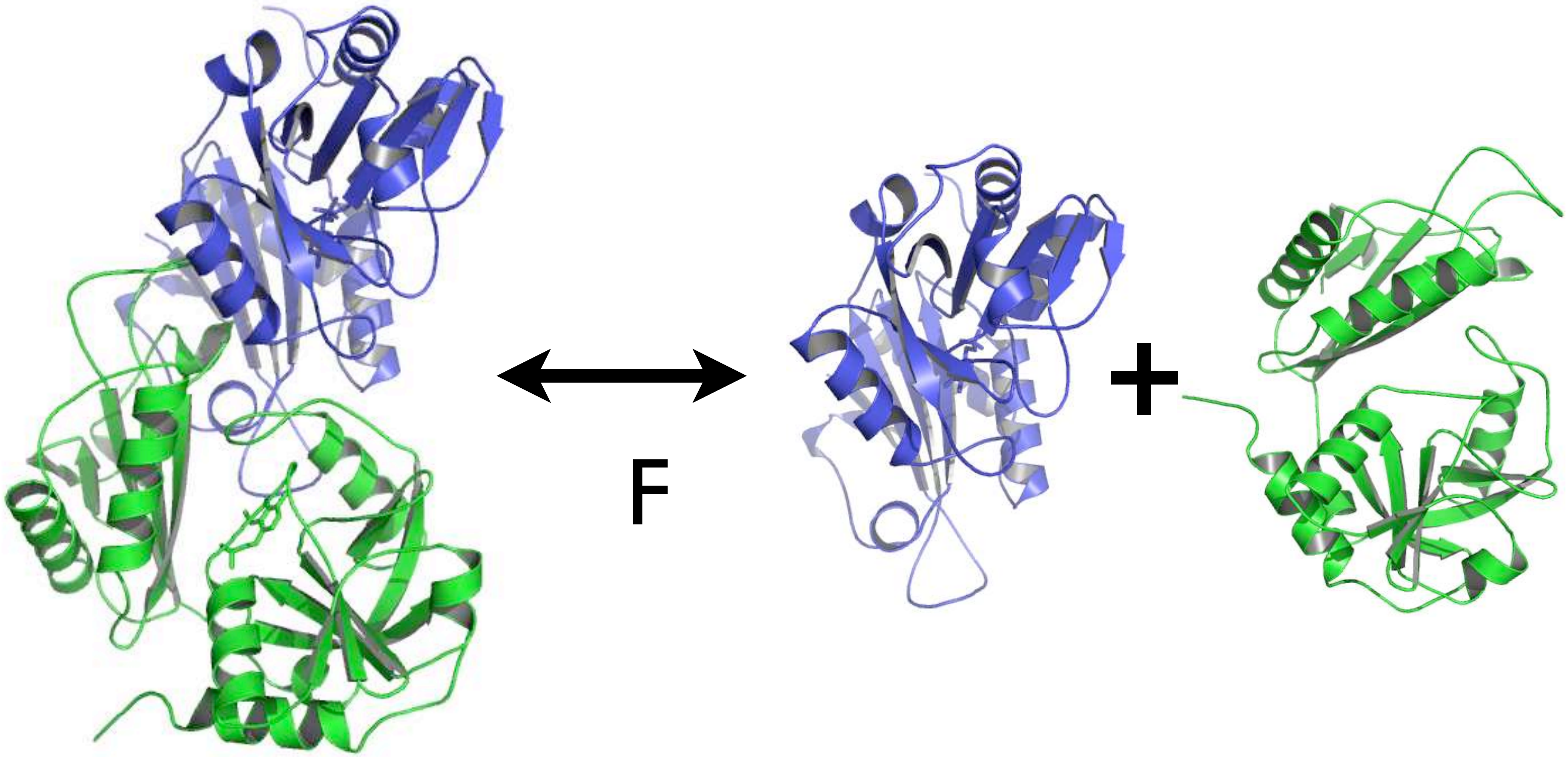
Quadratic Optimization to Predict Protein-Protein Interactions







- computational drug-design
- protein folding + *De Novo* design
- protein docking

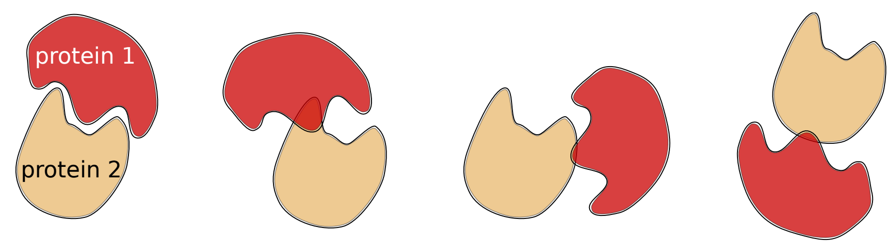
Sergei Grudinin

NANO-D INRIA Grenoble

Quadratic Optimization to Predict Protein-Protein Interactions

$$F(n(r)) \equiv F(n_{11}(r), .., n_{kl}(r), .., n_{MM}(r)) = \sum_{k=1}^{M} \sum_{l=k}^{M} \int_{0}^{r_{max}} n_{kl}(r) U_{kl}(r)\, dr$$

$$F(P_i^{nat}) < F(P_{ij}^{nonnat})$$



native   decoy #1   decoy #2   ...

Polynomial expansion of order *P*:

$$U_{kl}(r) = \sum_p w_p^{kl} \xi_p(r) \sqrt{\Omega(r)}, \qquad r \in [r_1; r_2]$$

$$n_{kl}(r) = \sum_p x_p^{kl} \xi_p(r) \sqrt{\Omega(r)}, \qquad r \in [r_1; r_2]$$

$$F(n(r)) \approx \sum_{k=1}^{M} \sum_{l=k}^{M} \sum_{p}^{P} w_p^{kl} x_p^{kl} = (\mathbf{w} \cdot \mathbf{x}),$$

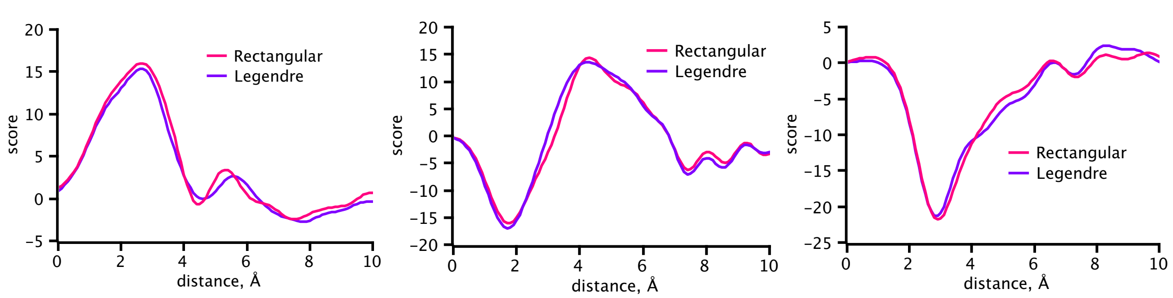
$$\mathbf{w}, \mathbf{x} \in \mathbb{R}^{P \times M \times (M+1)/2}$$

$$minimize: \quad \frac{\mathbf{w} \cdot \mathbf{w}}{2} + \sum_{i=0}^{m} C_{ij} \xi_{ij}$$

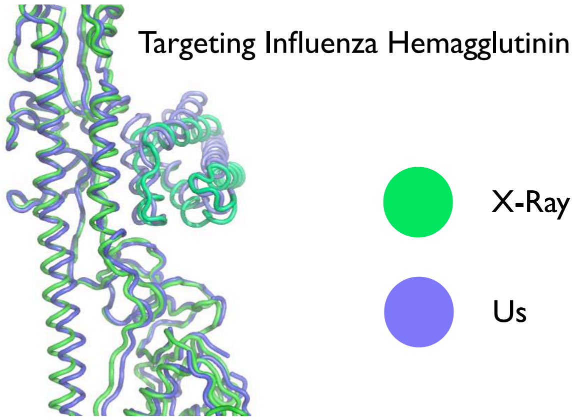
$$subject\ \ to: \quad y_{ij} \left[\mathbf{w} \cdot \mathbf{x}_{ij} + b \right] - 1 + \xi_{ij} \geq 0$$

$$\xi_{ij} \geq 0$$

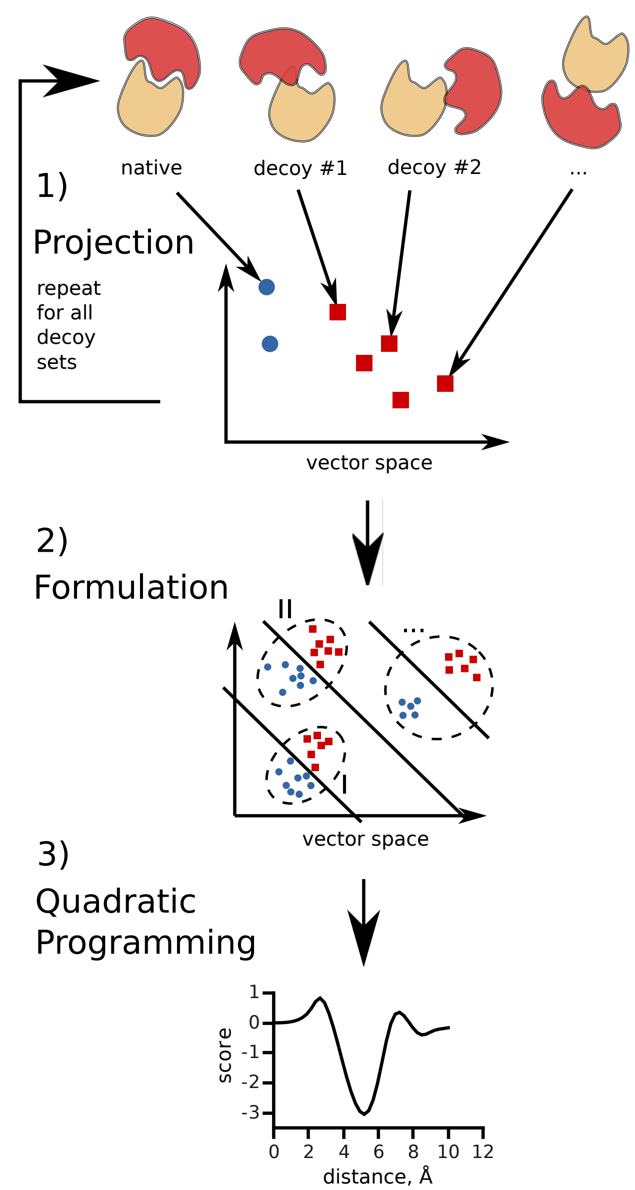
Energy Functions



Targeting Influenza Hemagglutinin



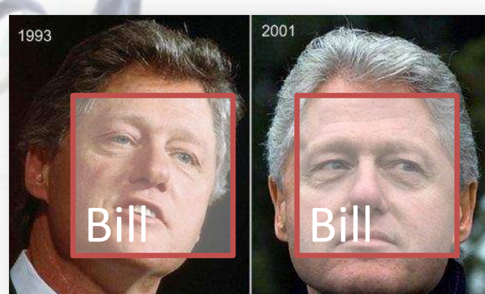
Algorithm



Sergei Grudinin   NANO-D INRIA Grenoble

I am watching you!

Inria
INVENTEURS DU MONDE NUMÉRIQUE
...SequeL...
Michal Valko



REAL-TIME LEARNING FROM STREAMS OF UNLABELED DATA

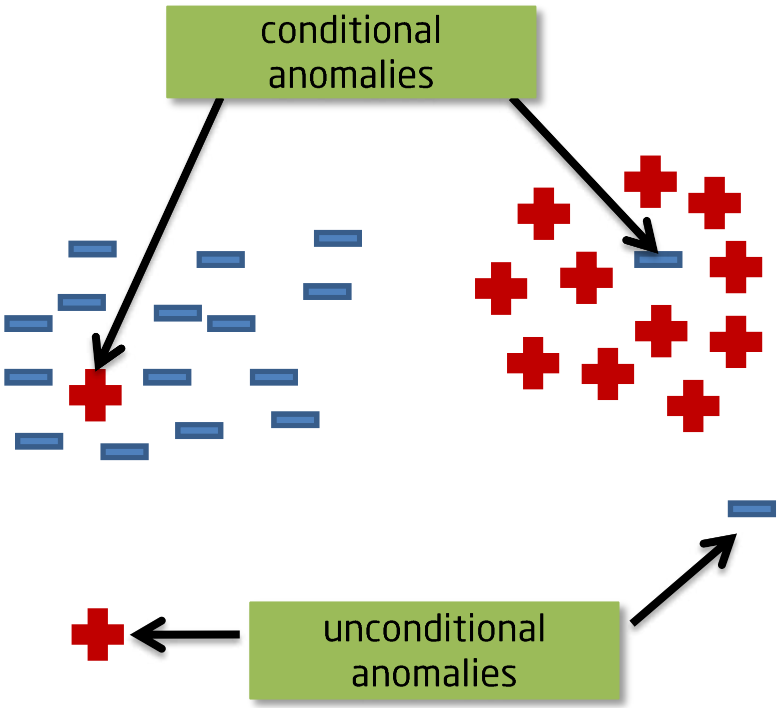
Features

- Minimal human feedback
- Adapts to changes in environment
- Robust to outliers
- Runs in real time
- High accuracy and recall
- Constant memory
- Bounds on performance

Online Semi-Supervised Learning on Quantized Graphs

Michal Valko

Conditional Anomaly Detection Using Soft Harmonic Functions: An Application to Clinical Alerting



- <u>Minimal human feedback</u>
- Problem: **Medical errors**
- Anomalies in **actions** (e.g. Med orders)
- Rule-based systems are costly
- Challenge: unconditional anomalies
- Estimate the **confidence** of actions
- Label propagation on data similarity graph
- Regularization to avoid overconfidence
- Results: Cardiac surgery patients (UPMC)

"***Medical errors account for 200 000 preventable deaths a year.***"

(HealthGrades study, Wall Street Journal, July 27th 2004)

"***Medical errors are the 8-th leading cause of death.***"

(To Err is Human: Building a safer health system., Kohn et. al, 2000)

www.scikit-learn.org
Machine Learning in Python

Some numbers

- 63,728 lines of Python code
- 10,371 lines of documentation
- 97 examples
- 32,400 entries on google.com
- 5,844 unique visitors in Oct. 2011 and 47,947 pageviews
- 38 people contributed to the latest 0.9 release
- 4983 emails on scikit-learn-general@lists.sf.net since Jan. 2010

Funding:

INVENTEURS DU MONDE NUMÉRIQUE

What to do with scikit-learn?

Supervised

- SVM (LibSVM, LibLinear)
- Lasso, Lars, Logistic
- Stoch. grad. descent
- Nearest Neighbors
- Partial Least Squares
- Naive Bayes
- Gaussian process
- Decision trees

Unsupervised

- Gaussian mixtures
- Manifold learning
- Mean-shift
- Affinity propagation
- K-Means
- Dictionary learning
- PCA / Kernel PCA
- NMF
- ICA

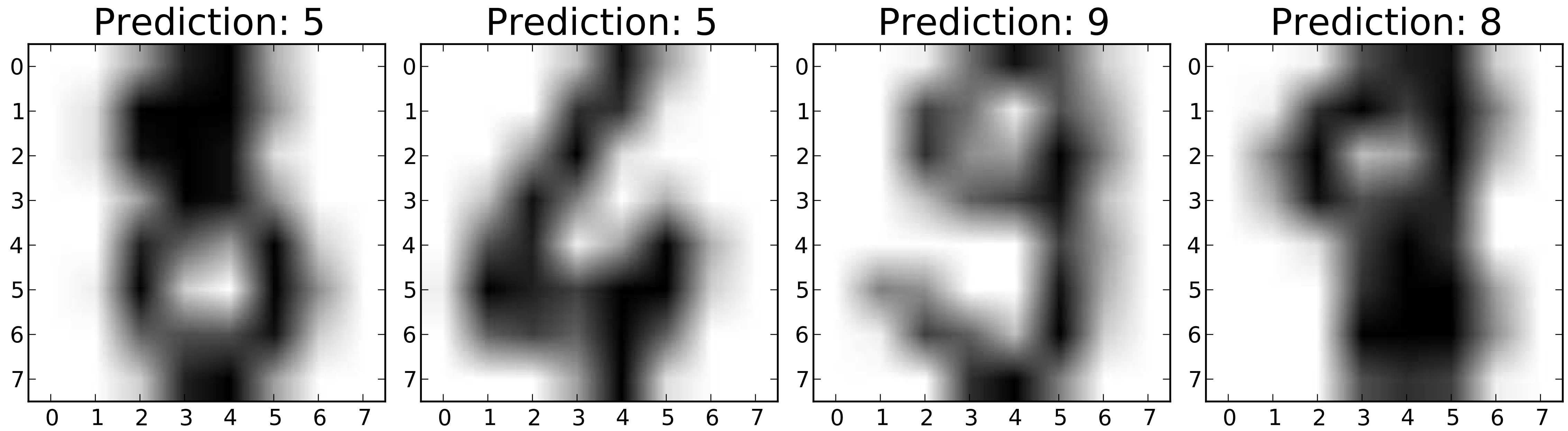
API

- Easy model selection with cross-validation
- Parallel computing for grid search of hyperparameters

Scikit-learn: Machine Learning in Python,
Pedregosa *et al., JMLR 12(Oct):2825−2830, 2011*.

Code: MNIST digits classif. with SVM

```python
import pylab as pl
from sklearn import datasets, svm
# Load data
digits = datasets.load_digits()
n_samples = len(digits.images)
data = digits.images.reshape((n_samples, -1))
# Learn
classifier = svm.SVC()
classifier.fit(data[:n_samples/2], digits.target[:n_samples/2])
# Predict and plot
for index, image in enumerate(digits.images[n_samples/2:n_samples/2+4]):
    pl.subplot(1, 4, index)
    pl.imshow(image, cmap=pl.cm.gray_r)
    pl.title('Prediction: %i' % classifier.predict(image.ravel()), fontsize=20)
```



A fast metric learning algorithm for kernel density estimation

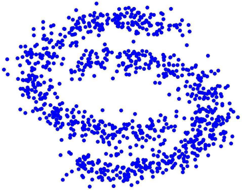
Nicolas Le Roux

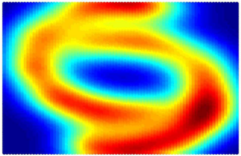
SIERRA

5/12/11

Kernel density estimation (KDE)

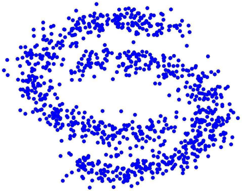
$$\mathbf{x}_1, \ldots, \mathbf{x}_N \sim p \qquad \widehat{p}(\mathbf{x}) = \frac{1}{N} \sum_j \mathcal{N}(\mathbf{x}, \mathbf{x}_j, \Sigma)$$

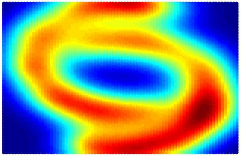




Kernel density estimation (KDE)

$$\mathbf{x}_1, \ldots, \mathbf{x}_N \sim p \qquad \widehat{p}(\mathbf{x}) = \frac{1}{N}\sum_j \mathcal{N}(\mathbf{x}, \mathbf{x}_j, \Sigma)$$

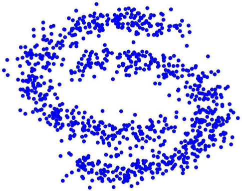


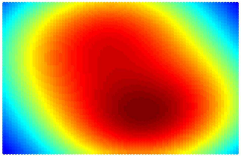


- Non-parametric
- Simple
- Heavily reliant on Σ
- Works poorly in high dimension

Kernel density estimation (KDE)

$$\mathbf{x}_1, \ldots, \mathbf{x}_N \sim p \qquad \widehat{p}(\mathbf{x}) = \frac{1}{N}\sum_j \mathcal{N}(\mathbf{x}, \mathbf{x}_j, \Sigma)$$





- Non-parametric
- Simple
- Heavily reliant on Σ
- Works poorly in high dimension

Why do we care ?

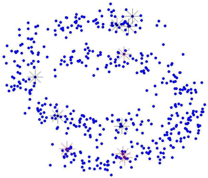
- Many methods rely on a distance measure (SSL, Spectral Clustering, k-NN)

Why do we care ?

- Many methods rely on a distance measure (SSL, Spectral Clustering, k-NN)
- Learning a good distance measure can improve performance

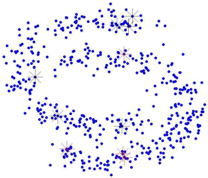
Why do we care ?

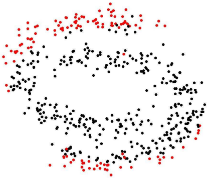
- Many methods rely on a distance measure (SSL, Spectral Clustering, k-NN)
- Learning a good distance measure can improve performance



Why do we care ?

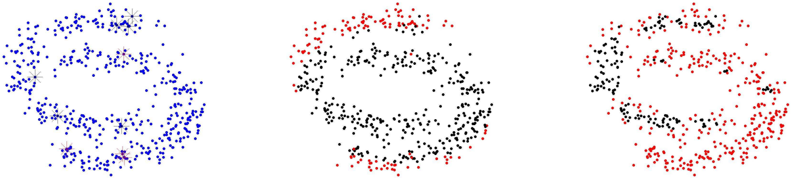
- Many methods rely on a distance measure (SSL, Spectral Clustering, k-NN)
- Learning a good distance measure can improve performance





Why do we care ?

- Many methods rely on a distance measure (SSL, Spectral Clustering, k-NN)
- Learning a good distance measure can improve performance



Local Component Analysis

- Automatic (and simple) selection of Σ
- Automatic (and simple) dimensionality reduction
- Fast and efficient approximation for large N.

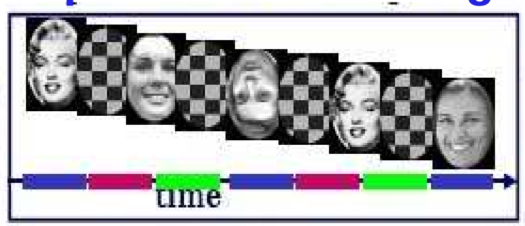
`http://nicolas.le-roux.name/code.html`

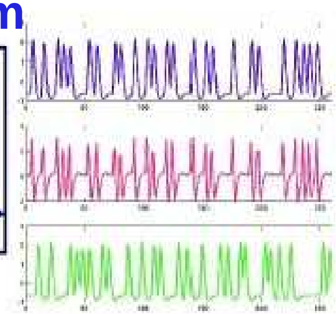
EXPERIMENTAL CONDITION SELECTION IN EVENT-RELATED FMRI

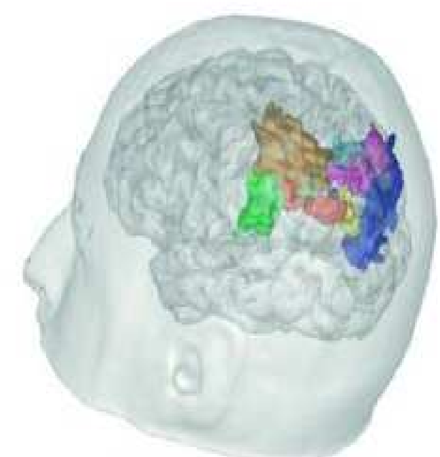
Christine Bakhous

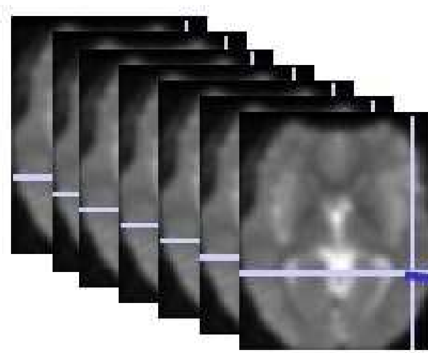
Supervisors :
F. Forbes, M. Dojat, P. Ciuciu

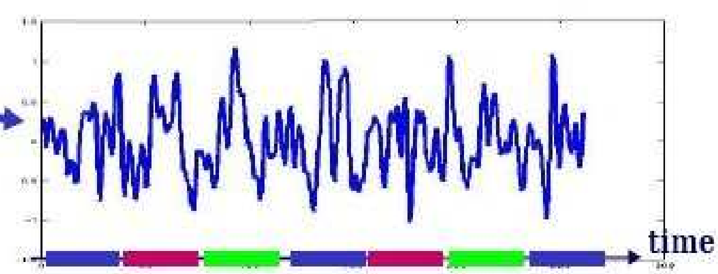
Experimental Paradigm







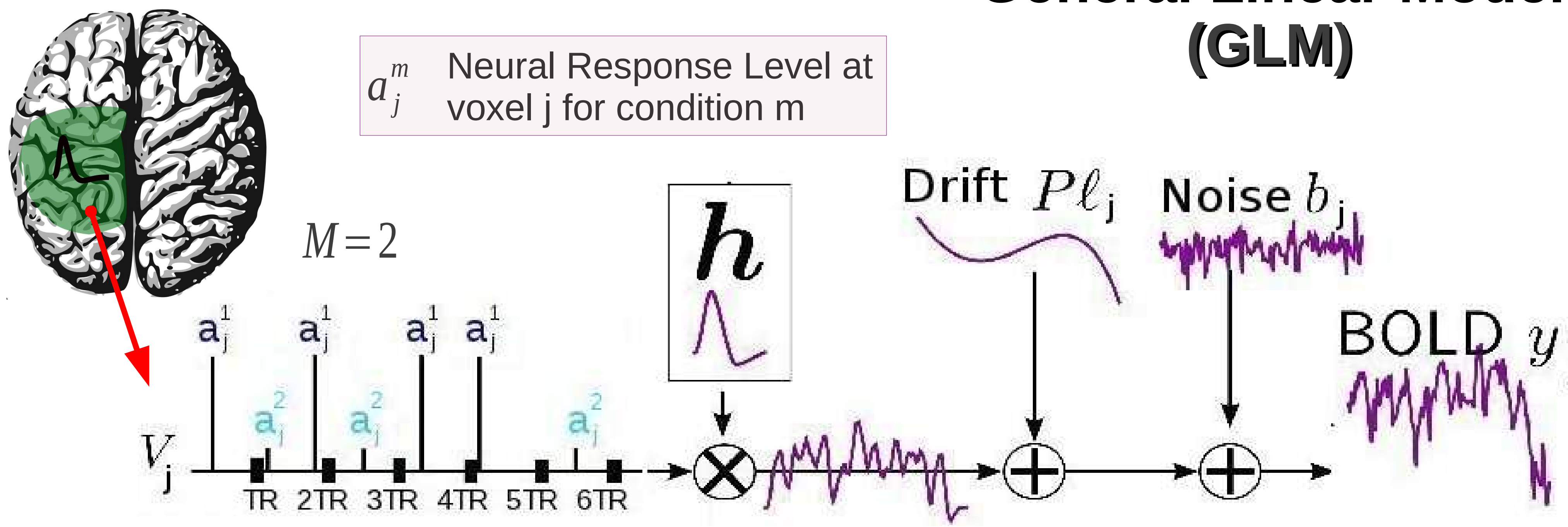




BOLD
Blood Oxygenation Level Dependent signal

79x95x46x128
3D signal + time

General Linear Model (GLM)



$$y_j = \sum_{m=1}^{M} a_j^m X^m h + P\ell_j + b_j$$

Complete Model

$$y_j = \sum_{m=1}^{M} a_j^m w^m X^m h + P\ell_j + b_j$$

Parsimonious Model

Selection of experimental conditions that best explain brain activity

Variable selection technique

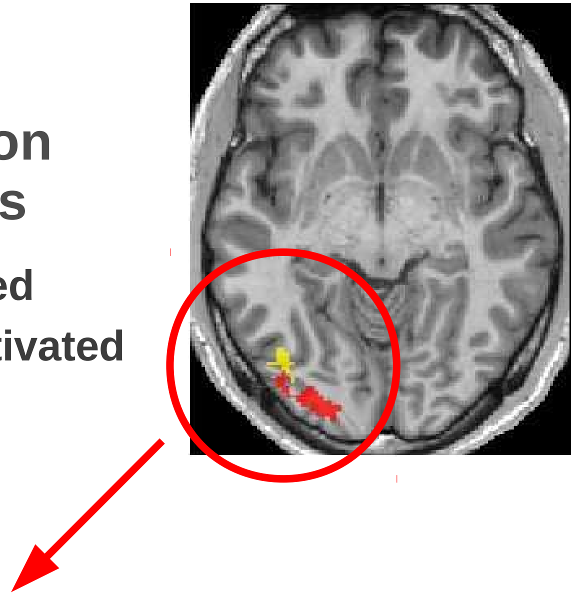
$p(w=1/y)$ Computed via ***Gibbs Sampler*** (***MCMC*** technique)

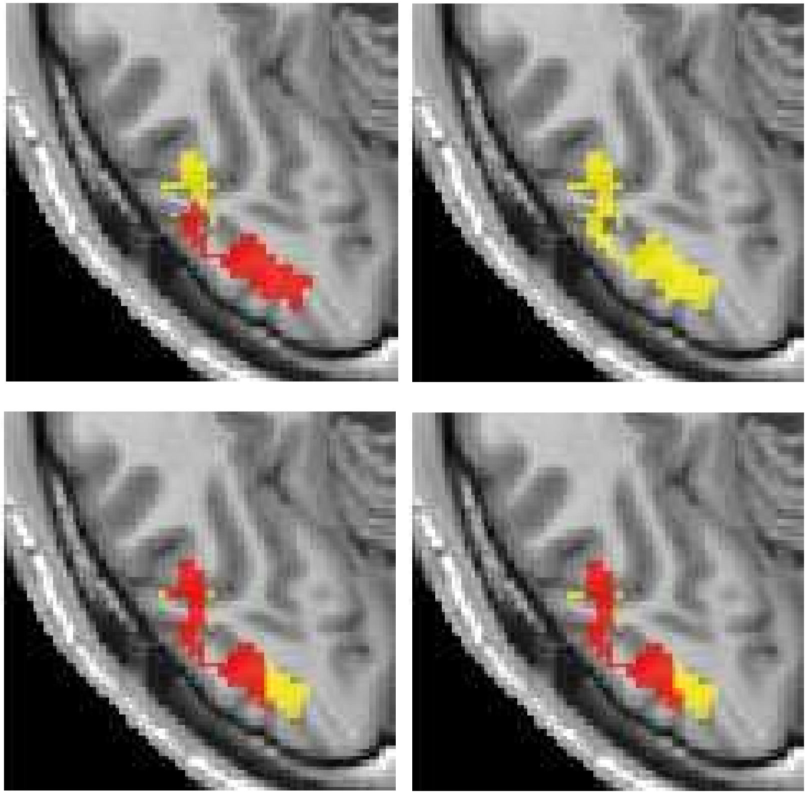
Experimental Data

ROI : visual left
Paradigm : **10** experimental conditions
6 visual (**relevant** in our ROI)
4 auditory (**irrelevant** in our ROI)

Activation classes

- Activated
- Non activated



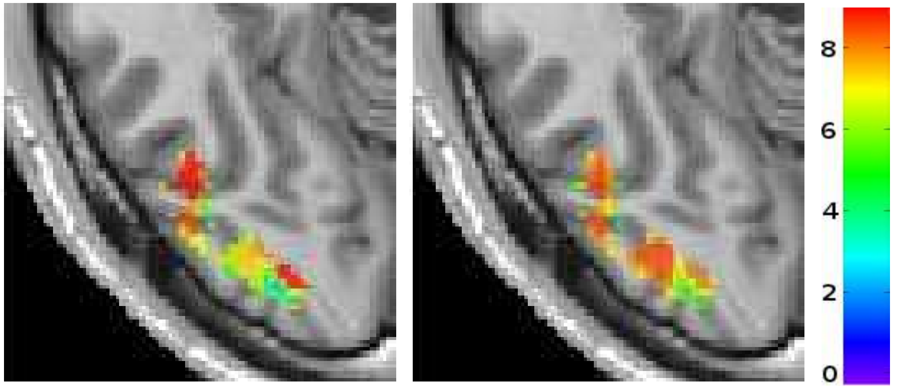


Auditory calculation

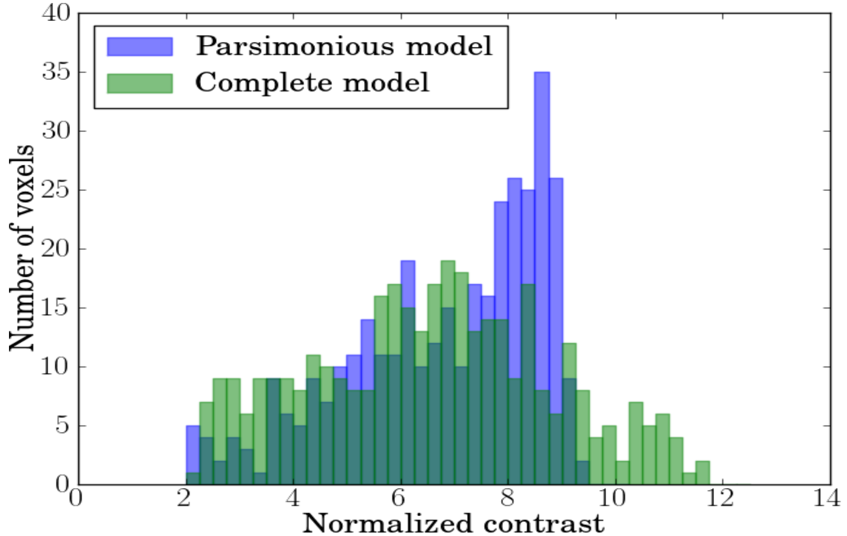
Vertical Checkerboard

Complete **Parsimonious**

Normalized contrast Video - Audio

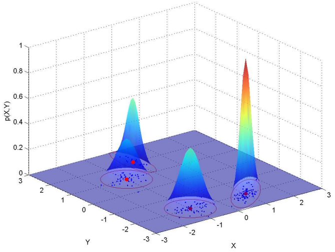


Complete **Parsimonious**



Bayesian Nonparametric Methods

- Learning of a model $f(\cdot|\theta)$ from data
- Bayesian nonparametrics :
  - Prior distribution on θ, $\dim(\theta) = \infty$
  - Allows the complexity of the model to increase as new data are gathered
  - Numerous applications : clustering, regression, latent feature extraction, etc.
  - Very active field of research in statistics and machine learning



- Contributions of the team ALEA
  - New BNP models for structured data
  - Efficient Monte Carlo algorithms to fit these models
  - Applications to dynamic clustering, ranking algorithms, extraction of features in images, etc.