A regularization algorithm for decoding perceptual profiles

Stefania Favilla[1], Marco Prato[2], Luca Zanni[2], Carlo A. Porro[1], Patrizia Baraldi[1]

1. Dipartimento di Scienze Biomediche, Università di Modena e Reggio Emilia, Italy

2. Dipartimento di Matematica, Università di Modena e Reggio Emilia, Italy

Introduction

We have tested and optimized a machine learning approach on fMRI data, based on new regularization learning algorithms [1] for regression problems. In this study, we have evaluated the feasibility of predicting the perceived pain intensity in healthy volunteers, from fMRI signals collected during an experimental model of acute prolonged (tonic) pain lasting several minutes [2,3].

Learning and regularization

Learning from examples: given a training set

$$\{(\boldsymbol{x}_i, y_i) \in X \times Y : i = 1, \dots, n\}, \; X \subseteq \mathbb{R}^m, \; Y \subseteq \mathbb{R}$$

find the decision function $f_\lambda : X \to Y$ to predict the label y of new examples **x** by solving [4]

$$\min_{f \in \mathcal{H}} \; (1/n) \sum_{i=1}^{n} V(y_i, f(\boldsymbol{x}_i)) + \lambda \|f\|_K^2 \qquad (1)$$

where: - V is a loss function
- $\mathcal{H}$ is a Reproducing Kernel Hilbert Space with Mercer kernel K [5]
- λ is a positive regularization parameter.

If we consider the quadratic loss $V(y, f(\boldsymbol{x})) = (y - f(\boldsymbol{x}))^2$ and the sampling operator $S_{\boldsymbol{x}} : \mathcal{H} \to \mathbb{R}^n$ defined by $(S_{\boldsymbol{x}} f)_i = f(\boldsymbol{x}_i)$, problem (1) becomes

$$\min_{f \in \mathcal{H}} \|S_{\boldsymbol{x}} f - \boldsymbol{y}\|_n^2 + \lambda \|f\|_K^2 \qquad (2)$$

where $\|\cdot\|_n$ is 1/n times the Euclidean norm in $\mathbb{R}^n$ and $\boldsymbol{y} = (y_1, \dots, y_n)^T$
The solution of (2) is the Tikhonov regularized solution of the linear inverse problem $S_{\boldsymbol{x}} f = \boldsymbol{y}$, whose explicit form is given by [6]

$$f_\lambda(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) \qquad (3)$$

Tikhonov regularization algorithm: $\boldsymbol{\alpha} = (\boldsymbol{K} + n\lambda I)^{-1} \boldsymbol{y}$, where $\boldsymbol{K}_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j), \; \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$

ν – method [7]: $\begin{cases} \boldsymbol{\alpha}^{(i)} = \boldsymbol{\alpha}^{(i-1)} + u_i(\boldsymbol{\alpha}^{(i-1)} - \boldsymbol{\alpha}^{(i-2)}) + \frac{\omega_i}{n}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\alpha}^{(i-1)}) \\ \boldsymbol{\alpha}^{(0)} = \boldsymbol{0} \qquad i = 1, 2, \dots, t \end{cases}$

$$u_i = \frac{(i-1)(2i-3)(2i+2\nu-1)}{(i+2\nu-1)(2i+4\nu-1)(2i+2\nu-3)}, \quad \omega_i = 4\frac{(2i+2\nu-1)(i+\nu-1)}{(i+2\nu-1)(2i+4\nu-1)}$$

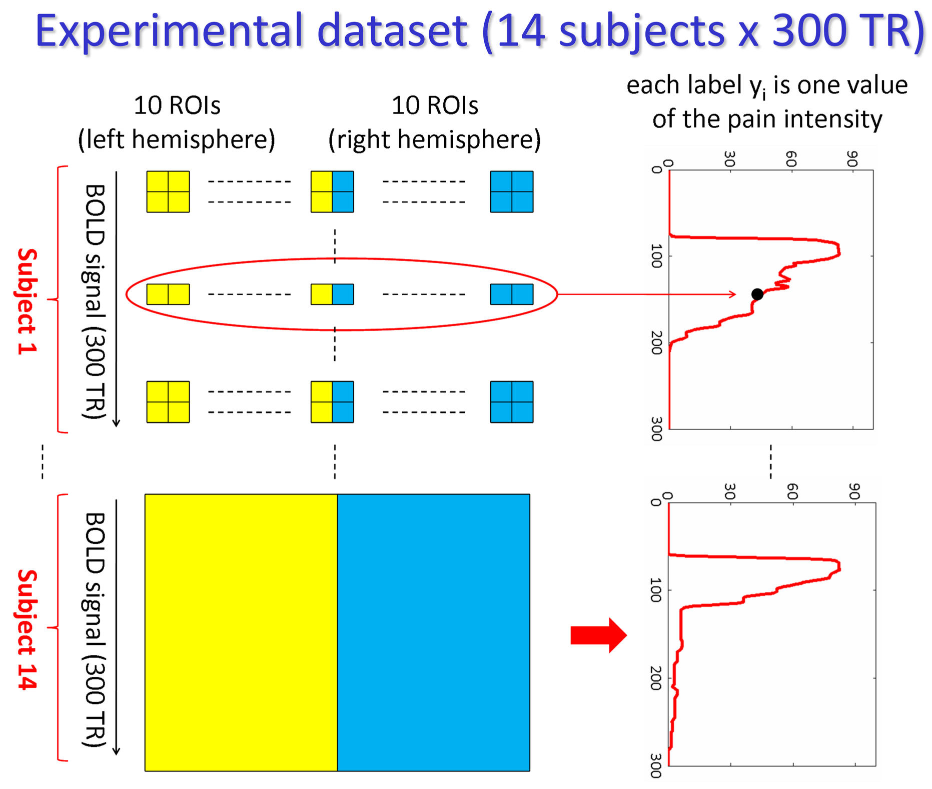
<u>Crucial parameter</u>: the role of the regularization parameter λ is played by the number of iterations t. The fast convergence together with the very simple definition of each iteration (which involves only a matrix-vector product) makes the ν-method less computationally expensive than state-of-the-art approaches as SVM [8].

Data pre-processing and the input features

➢ Corrections of head movements, temporal filtering (gaussian 15 points, σ=100) and linear detrending

➢ **Features:** for each ROI the first principal singular vector (1st-SVD) of the time series was computed using the AFNI S/W package. To roughly recover the original signal variation, each 1st-SVD was multiplied by the median of the standard deviations of the time courses of each ROI voxel, computed over an interval of 6.5 min, starting from 30s before stimulation

Experimental dataset (14 subjects x 300 TR)



❖ **Training set** (13 volunteers x 300 TR)
❖ **Model selection**: linear kernel

$$K(\boldsymbol{x}_j, \boldsymbol{x}_i) = \boldsymbol{x}_j^T \boldsymbol{x}_i,$$

❖ **Learning algorithm**: ν – method (ν=5)

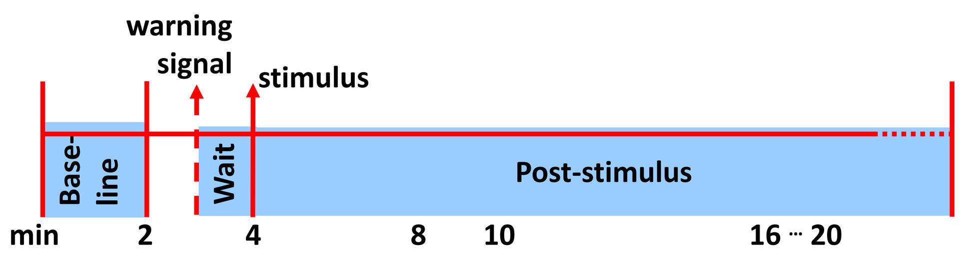
Test set
(1 volunteer x 300 TR)

<u>Parameters' choice</u>
the number of iterations t is chosen via a "**leave-one-volunteer-out**" cross validation

The fMRI experiment

➢**14 volunteers** were injected subcutaneously with a dilute ascorbic acid solution into the thenar eminence of the left hand

Experimental design



➢ **pain intensity** was recorded by moving a computer-controlled visual analogue scale (VAS in 0-100) with the right (unstimulated) hand

➢ **functional images** were acquired by a GE 1.5T scanner, using an EPI BOLD-sensitive sequence (TR= 4s, 3.75x3.75x4mm interpolated to 2x2x2mm)

➢ **300 brain volumes** (in 20 min) were collected from 24 contiguous axial planes covering the diencephalon and telencephalon

➢ **10 regions of interest** (ROIs) -based on *a priori* hypothesis- were identified in both hemispheres on anatomical brain images, acquired in the same experimental session. They included the thalamus, basal ganglia, parietal operculum, insular and cingulate cortical areas.

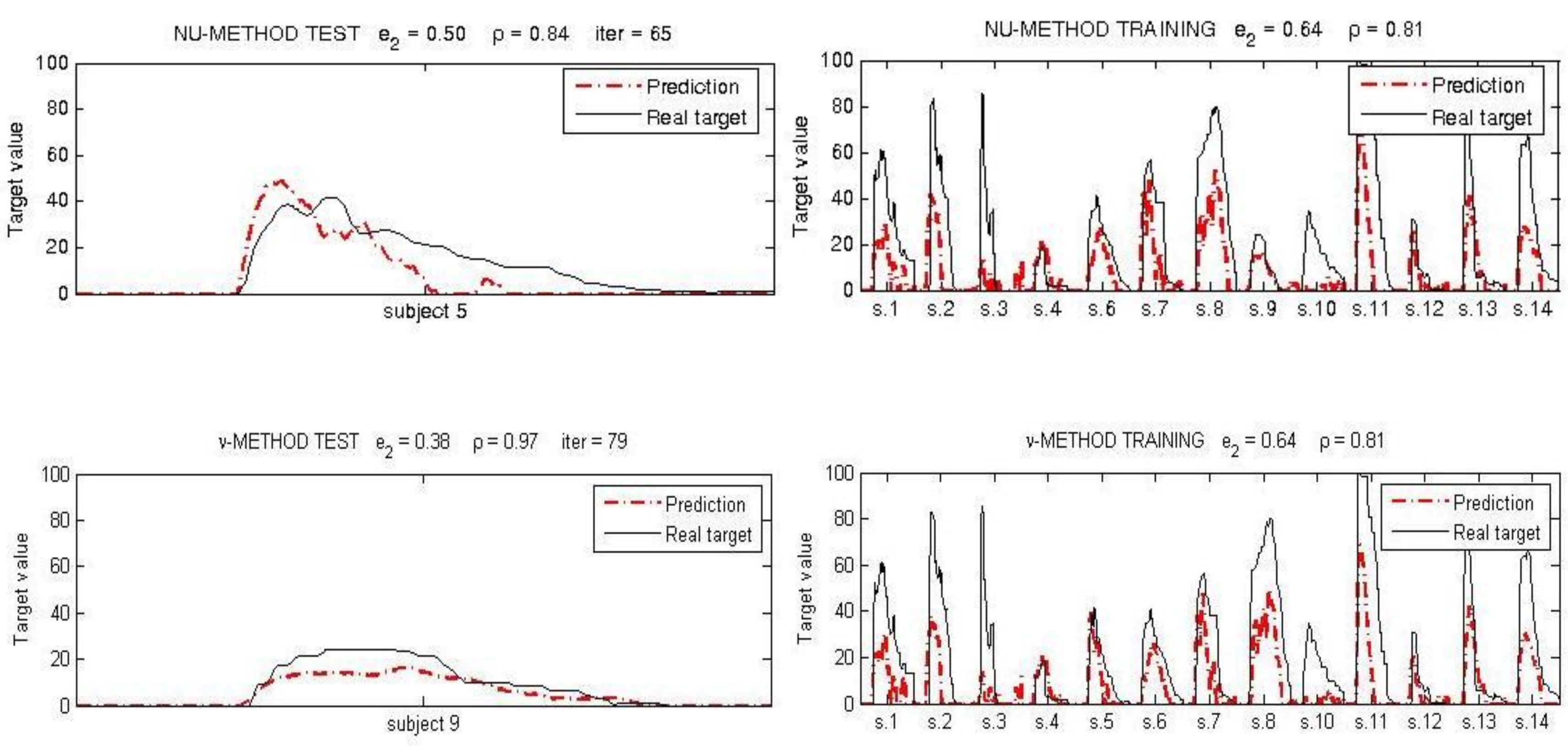
Results

Performance evaluation

- relative reconstruction error $\rho_2(\boldsymbol{f_\lambda}, \boldsymbol{y}) = \|\boldsymbol{f_\lambda} - \boldsymbol{y}\|_2 / \|\boldsymbol{y}\|_2$
- Pearson cross-correlation coefficient $\rho_P(\boldsymbol{f_\lambda}, \boldsymbol{y})$

Predicted pain profile compared to the real target for two representative subjects

TEST TRAINING



	Performance on Tests													
ρ_2	0.73	0.74	0.92	0.50	0.50	0.53	0.66	0.68	**0.38**	0.80	0.77	0.58	0.67	0.83
ρ_P	0.87	0.71	0.44	0.88	0.84	0.90	0.68	0.82	**0.97**	0.81	0.81	0.79	0.70	0.83
Subj.	1	2	3	4	5	6	7	8	**9**	10	11	12	13	14

CONCLUSIONS

Albeit preliminary, these results show that the proposed algorithm is potentially useful for inter-subject prediction of perceptual profiles from fMRI data.

Given its fast convergence rate, the ν-method is suitable for regression or multiple classification problems in the fMRI field.

[1] Lo Gerfo L., et al. (2008), Spectral algorithms for supervised learning, *Neural Computation* **20**, 1873-1897
[2] Porro C.A., et al. (1998), Temporal and intensity coding of pain in human cortex, *Journal of Neurophysiology* **80**, 3312-3320
[3] Baraldi P., et al. (2003), Perceptual coding of pain intensity: where in the brain? 9th HBM 2003, New York
[4] Vapnik V.N. (1998), *Statistical learning theory*. New York: John Wiley and Sons
[5] Cucker F., Smale S. (2002), On the mathematical foundations of learning, *Bulletin of the American Mathematical Society* **39**, 1-49
[6] Tikhonov A.N., Arsenin V.Y. (1977), *Solution of Ill Posed Problems*, Washington DC: WH Winston
[7] Engl H.W., Hanke M., Neubauer A. (1996), *Regularization of Inverse Problems*, Dordrecht: Kluwer Academic Publishers Group
[8] Cristianini N., Shawe-Taylor J. (2000), *Support Vector Machines and other kernel-based learning methods.* Cambridge: Cambridge University Press