Projections to Latent Space

Roman Isachenko

Advisor: Vadim Strijov

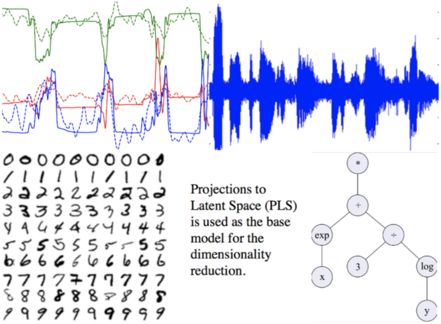
September 26, 2018

Complex structured objects

There is no explicit feature description.

Common structured data

- time series
- speech
- images
- text
- graphs

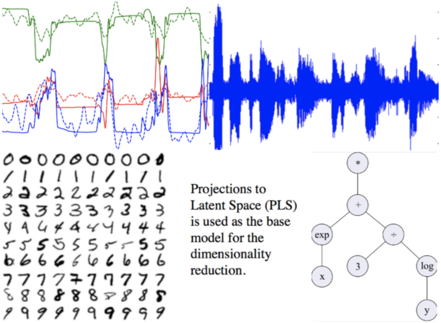


Complex structured objects

There is no explicit feature description.

Common structured data

- time series
- speech
- images
- text
- graphs



Solution

- local models
- expert-defined functions
- probabilistic models
- sparse models
- representation learning

Dependencies

- input space
- target space

PCA, CCA, PLS

Principal Component Analysis

$$\max_{\mathbf{w}} \|\mathbf{X}\mathbf{w}\|_2, \quad \text{s.t. } \|\mathbf{w}\|_2 = 1, \quad \max_{\mathbf{c}} \|\mathbf{Y}\mathbf{c}\|_2, \quad \text{s.t. } \|\mathbf{c}\|_2 = 1$$

Canonical Correlation Analysis

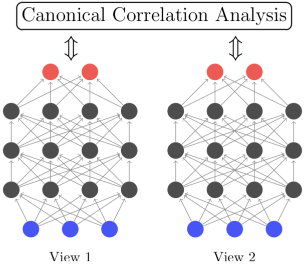
$$\max_{\mathbf{w},\mathbf{c}} \operatorname{corr}(\mathbf{Y}\mathbf{c}, \mathbf{X}\mathbf{w}), \quad \text{s.t. } \|\mathbf{w}\|_2 = \|\mathbf{c}\|_2 = 1$$

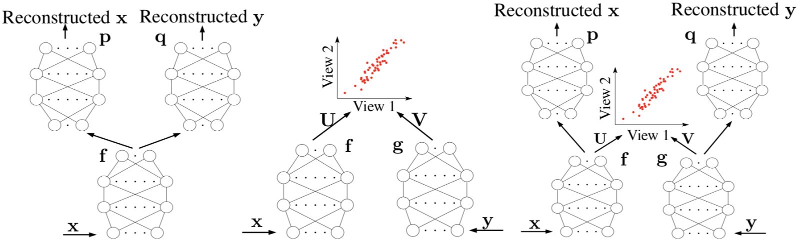
Partial Least Squares

$$\max_{\mathbf{w},\mathbf{c}} \operatorname{cov}(\mathbf{Y}\mathbf{c}, \mathbf{X}\mathbf{w}), \quad \text{s.t. } \|\mathbf{w}\|_2 = \|\mathbf{c}\|_2 = 1$$

$$\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \operatorname{corr}(\mathbf{x}, \mathbf{y}) \cdot \operatorname{var}(\mathbf{x}) \cdot \operatorname{var}(\mathbf{y})$$

Nonlinear PLS





(a) SplitAE (b) DCCA (c) DCCAE/CorrAE/DistAE

Applications

Multi-view data

- multilingual translation
- images from different domains
- speech processing
- multimodal signal processing

Literature

- Andrew G. et al. Deep canonical correlation analysis // International Conference on Machine Learning. – 2013.
- Wang W. et al. On deep multi-view representation learning // International Conference on Machine Learning. – 2015.
- Tran L., Yin X., Liu X. Disentangled representation learning gan for pose-invariant face recognition //CVPR. – 2017.

Signal decoding problem

Goal

Investigate dependencies in both input and target spaces and build a stable model for signal decoding in the case of multicorrelated object description.

Challenge

- Measurements of spatial-temporal data are interacted since sensors are close to each other.
- Target variable is a vector whose elements are dependent.

Solution

Propose dimensionality reduction and feature selection algorithms which take into account dependencies in both input and target spaces.

Related works

- Katrutsa A., Strijov V. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria // *Expert Systems with Applications* 76, 2017.
- Li J. et al. Feature selection: A data perspective //*ACM Computing Surveys (CSUR)* 50(6), 2017.
- Eliseyev A. et al. Iterative N-way partial least squares for a binary self-paced brain–computer interface in freely moving animals //*Journal of neural engineering* 4(8), 2011.
- Rodriguez-Lujan I. et al. Quadratic programming feature selection // *Journal of Machine Learning Research* 11(Apr), 2010.
- Motrenko A., Strijov V. Multi-way Feature Selection for ECoG-based Brain-Computer Interface // *Expert Systems with Applications* Submitted to the journal.

Multivariate regression

Given

Dataset $(\mathbf{X}, \mathbf{Y})$, design matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, target matrix $\mathbf{Y} \in \mathbb{R}^{m \times r}$,

$$\mathbf{X} = [\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_n]; \quad \mathbf{Y} = [\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_r].$$

Model

Forecast a dependent variable $\mathbf{y} \in \mathbb{R}^r$ from an independent input object $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{y} = \boldsymbol{\Theta}\mathbf{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\Theta} \in \mathbb{R}^{r \times n}.$$

Loss function

$$\mathcal{L}(\boldsymbol{\Theta}|\mathbf{X}, \mathbf{Y}) = \left\| \underset{m \times r}{\mathbf{Y}} - \underset{m \times n}{\mathbf{X}} \cdot \underset{r \times n}{\boldsymbol{\Theta}}^{\mathsf{T}} \right\|_2^2 \to \min_{\boldsymbol{\Theta}}.$$

$$\boldsymbol{\Theta}^{\mathsf{T}} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{Y}.$$

The linear dependent columns of the matrix $\mathbf{X}$ leads to an instable solution. To avoid the strong linear dependence, dimensionality reduction and feature selection techniques are used.

Goal

- project the matrices $\mathbf{X}$ and $\mathbf{Y}$ into joint latent space;
- maximize covariance between the projections;
- save variance of the initial matrices.

Partial Least Squares (PLS) regression

$$\underset{m\times n}{\mathbf{X}} = \underset{m\times l}{\mathbf{T}} \cdot \underset{l\times n}{\mathbf{P}^{\mathsf{T}}} + \underset{m\times n}{\mathbf{F}} = \sum_{k=1}^{l} \underset{m\times 1}{\mathbf{t}_k} \cdot \underset{1\times n}{\mathbf{p}_k^{\mathsf{T}}} + \underset{m\times n}{\mathbf{F}},$$

$$\underset{m\times r}{\mathbf{Y}} = \underset{m\times l}{\mathbf{U}} \cdot \underset{l\times r}{\mathbf{Q}^{\mathsf{T}}} + \underset{m\times r}{\mathbf{E}} = \sum_{k=1}^{l} \underset{m\times 1}{\mathbf{u}_k} \cdot \underset{1\times r}{\mathbf{q}_k^{\mathsf{T}}} + \underset{m\times r}{\mathbf{E}}.$$

$$\mathbf{U} \approx \mathbf{TB}, \quad \mathbf{B} = \mathrm{diag}(\beta_k), \quad \beta_k = \mathbf{u}_k^{\mathsf{T}}\mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}}\mathbf{t}_k).$$

PLS pseudocode

Require: $\mathbf{X}, \mathbf{Y}, l$;
Ensure: $\mathbf{T}, \mathbf{P}, \mathbf{Q}$;

normalize matrices $\mathbf{X}$ и $\mathbf{Y}$ by columns
initialize $\mathbf{u}_0$ (the first column of $\mathbf{Y}$)
$\mathbf{X}_1 = \mathbf{X}; \mathbf{Y}_1 = \mathbf{Y}$
for $k = 1, \dots, l$ **do**
  repeat
    $\mathbf{w}_k := \mathbf{X}_k^{\mathsf{T}} \mathbf{u}_{k-1} / (\mathbf{u}_{k-1}^{\mathsf{T}} \mathbf{u}_{k-1}); \quad \mathbf{w}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$
    $\mathbf{t}_k := \mathbf{X}_k \mathbf{w}_k$
    $\mathbf{c}_k := \mathbf{Y}_k^{\mathsf{T}} \mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}} \mathbf{t}_k); \quad \mathbf{c}_k := \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|}$
    $\mathbf{u}_k := \mathbf{Y}_k \mathbf{c}_k$
  until $\mathbf{t}_k$ stabilizes
  $\mathbf{p}_k := \mathbf{X}_k^{\mathsf{T}} \mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}} \mathbf{t}_k), \ \mathbf{q}_k := \mathbf{Y}_k^{\mathsf{T}} \mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}} \mathbf{t}_k)$
  $\mathbf{X}_{k+1} := \mathbf{X}_k - \mathbf{t}_k \mathbf{p}_k^{\mathsf{T}}$
  $\mathbf{Y}_{k+1} := \mathbf{Y}_k - \mathbf{t}_k \mathbf{q}_k^{\mathsf{T}}$

PLS regression

Statement (Isachenko, 2017)

The best description of the matrices $\mathbf{X}$ *and* $\mathbf{Y}$ *taking into account their interrelation is achieved by maximization of the covariance between the vectors* $\mathbf{t}_k$ *and* $\mathbf{u}_k$*.*

Statement (Isachenko, 2017)

The vector $\mathbf{w}_k$ *and* $\mathbf{c}_k$ *are eigenvectors of the matrices* $\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k$ *and* $\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k$*, corresponding to the maximum eigenvalues.*

Statement (Isachenko, 2017)

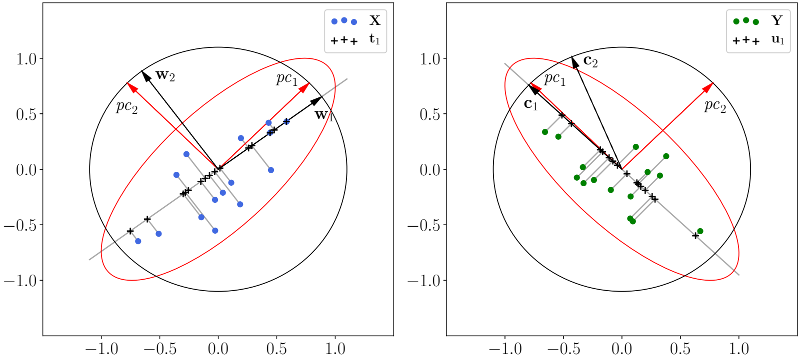
The update rule for the vectors in steps (6)–(9) corresponds to the maximization of the covariance between the vectors $\mathbf{t}_k$ *and* $\mathbf{u}_k$*.*

PLS regression model

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} \approx \mathbf{T}\mathbf{B}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} = \mathbf{X}\mathbf{W}^{*}\mathbf{B}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} = \mathbf{X}\boldsymbol{\Theta} + \mathbf{E}.$$

$$\boldsymbol{\Theta} = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{B}\mathbf{Q}^{\mathsf{T}}, \quad \mathbf{T} = \mathbf{X}\mathbf{W}^{*}, \quad \text{where } \mathbf{W}^{*} = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}.$$

PLS example for two-dimensional case



Feature selection problem

Goal

Find a boolean vector $\mathbf{a} = \{0, 1\}^n$ of indicators for selected features.

Feature selection error function

$$\mathbf{a} = \arg\min_{\mathbf{a}' \in \{0,1\}^n} S(\mathbf{a}'|\mathbf{X}, \mathbf{Y}).$$

Relaxed problem

From discrete domain $\{0, 1\}^n$ to continuous relaxation $[0, 1]^n$:

$$\mathbf{z} = \arg\min_{\mathbf{z}' \in [0,1]^n} S(\mathbf{z}'|\mathbf{X}, \mathbf{Y}), \quad a_j = [z_j > \tau].$$

Once the solution $\mathbf{a}$ is known:

$$\mathcal{L}(\boldsymbol{\Theta}_\mathbf{a}|\mathbf{X}_\mathbf{a}, \mathbf{Y}) = \left\|\mathbf{Y} - \mathbf{X}_\mathbf{a}\boldsymbol{\Theta}_\mathbf{a}^\mathsf{T}\right\|_2^2 \to \min_{\boldsymbol{\Theta}_\mathbf{a}},$$

where the subscript $\mathbf{a}$ indicates the submatrix with the columns for which $a_j = 1$.

Quadratic Programming Feature Selection

$$\|\boldsymbol{\nu} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \to \min_{\boldsymbol{\theta} \in \mathbb{R}^n}.$$

Quadratic programming problem

$$S(\mathbf{z}|\mathbf{X}, \boldsymbol{\nu}) = (1-\alpha) \cdot \underbrace{\mathbf{z}^{\mathsf{T}}\mathbf{Q}\mathbf{z}}_{\mathrm{Sim}(\mathbf{X})} - \alpha \cdot \underbrace{\mathbf{b}^{\mathsf{T}}\mathbf{z}}_{\mathrm{Rel}(\mathbf{X}, \boldsymbol{\nu})} \to \min_{\substack{\mathbf{z} \geq \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}}\mathbf{z} = 1}}.$$

- $\mathbf{z} \in [0, 1]^n$ – feature importances;
- $\mathbf{Q} \in \mathbb{R}^{n \times n}$ – pairwise feature similarities;
- $\mathbf{b} \in \mathbb{R}^n$ – feature relevances to the target vector.

$$\mathbf{Q} = \left[|\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)|\right]_{i,j=1}^n, \quad \mathbf{b} = \left[|\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu})|\right]_{i=1}^n.$$

Statement

In the case of semidefinite matrix $\mathbf{Q}$ the QPFS problem is convex. Shift spectrum for semidefinite relaxation:

$$\mathbf{Q} \to \mathbf{Q} - \lambda_{\min}\mathbf{I}.$$

Relevance Aggregation (RelAgg)

$$\mathbf{b} = [|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu})|]_{i=1}^{n} \to \mathbf{b} = \left[\sum_{k=1}^{r} |\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_k)|\right]_{i=1}^{n}.$$

Drawback: the approach does not use the dependencies in the columns of $\mathbf{Y}$.

Symmetric Importances (SymImp)

Penalize correlated targets by $\text{Sim}(\mathbf{Y})$

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\text{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\text{Rel}(\mathbf{X},\mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\text{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n, \, \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1 \\ \mathbf{z}_y \geq \mathbf{0}_r, \, \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

$$\mathbf{Q}_x = \left[|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)|\right]_{i,j=1}^{n}, \; \mathbf{Q}_y = [|\text{corr}(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j)|]_{i,j=1}^{r}, \; \mathbf{B} = [|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j)|]_{\substack{i=1,\dots,n \\ j=1,\dots,r}}.$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1, \quad \alpha_i \geq 0.$$

Multivariate QPFS

SymImp penalizes targets that are correlated and are explained by features to a lesser extent.

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}}; \quad \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} + \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

Minimax approach (MinMax)

$$\min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \left(\text{or} \max_{\substack{\mathbf{z}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \right) \left[\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} - \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \right].$$

Theorem (Isachenko, 2018)

For positive definite matrices $\mathbf{Q}_x$ and $\mathbf{Q}_y$ the maxmin and minmax problems have the same optimal value.

Theorem (Isachenko, 2018)

Minimax problem is equivalent to the quadratic problem with $n + r + 1$ variables.

Shift spectrum to obtain the convex semidefinite relaxation.

Multivariate QPFS

Maximum Relevances (MaxRel)

$$\min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^\mathsf{T}\mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^\mathsf{T}\mathbf{z}_y = 1}} \left[(1-\alpha) \cdot \mathbf{z}_x^\mathsf{T} \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^\mathsf{T} \mathbf{B} \mathbf{z}_y \right].$$

Theorem (Isachenko, 2018)

For positive definite matrices $\mathbf{Q}_x$ the max-min and min-max problems have the same optimal value and the final quadratic problem is convex.

Assymmetric importances (AsymImp)

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^\mathsf{T} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\left(\mathbf{z}_x^\mathsf{T} \mathbf{B} \mathbf{z}_y - \mathbf{b}^\mathsf{T} \mathbf{z}_y \right)}_{\mathrm{Rel}(\mathbf{X}, \mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^\mathsf{T} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n,\, \mathbf{1}_n^\mathsf{T}\mathbf{z}_x = 1 \\ \mathbf{z}_y \geq \mathbf{0}_r,\, \mathbf{1}_r^\mathsf{T}\mathbf{z}_y = 1}}.$$

For $b_j = \max\limits_{i=1,\dots,n} [\mathbf{B}]_{i,j}$ coefficients of $\mathbf{z}_y$ in $\mathrm{Rel}(\mathbf{X}, \mathbf{Y})$ are non-negative.

Statement

For the univariate case $r = 1$ the proposed strategies SymImp, MinMax, MaxRel, AsymImp coincide with the original QPFS algorithm.

Summary of the proposed algorithms

Algorithm	Problem	Error function $S(\mathbf{z}\vert\mathbf{X},\mathbf{Y})$
RelAgg	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$	$\min_{\mathbf{z}_x}\left[(1-\alpha)\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{1}_r\right]$
SymImp	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\right]$	$\min_{\mathbf{z}_x,\,\mathbf{z}_y}\left[\alpha_1\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha_2\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y + \alpha_3\cdot\mathbf{z}_y^{\mathsf{T}}\mathbf{Q}_y\mathbf{z}_y\right]$
MinMax	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$ $\max\left[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\right]$	$\min_{\mathbf{z}_x}\max_{\mathbf{z}_y}\left[\alpha_1\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha_2\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y - \alpha_3\cdot\mathbf{z}_y^{\mathsf{T}}\mathbf{Q}_y\mathbf{z}_y\right]$
MaxRel	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$ $\max\left[\mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$	$\min_{\mathbf{z}_x}\max_{\mathbf{z}_y}\left[(1-\alpha)\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y\right]$
AsymImp	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$ $\max\left[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\right]$	$\min_{\mathbf{z}_x,\mathbf{z}_y}\left[\alpha_1\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha_2\left(\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y - \mathbf{b}^{\mathsf{T}}\mathbf{z}_y\right) + \alpha_3\mathbf{z}_y^{\mathsf{T}}\mathbf{Q}_y\mathbf{z}_y\right]$

Quality criteria

Scaled RMSE

Prediction quality:

$$\text{sRMSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathrm{a}) = \sqrt{\frac{\text{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathrm{a})}{\text{MSE}(\mathbf{Y}, \overline{\mathbf{Y}})}} = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}_\mathrm{a}\|_2}{\|\mathbf{Y} - \overline{\mathbf{Y}}\|_2}, \quad \text{where} \quad \widehat{\mathbf{Y}}_\mathrm{a} = \mathbf{X}_\mathrm{a}\boldsymbol{\Theta}_\mathrm{a}^\mathsf{T}.$$

$\overline{\mathbf{Y}}$ is a constant prediction.

Multicorrelation

Mean value of miltiple correlation coefficient:

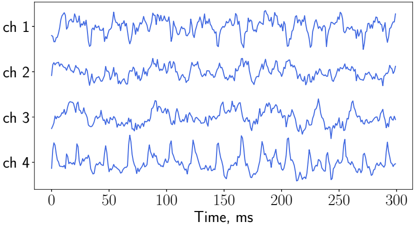
$$R^2 = \frac{1}{r}\text{tr}\left(\mathbf{C}^\mathsf{T}\mathbf{R}^{-1}\mathbf{C}\right); \quad \mathbf{C} = [\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j)]_{\substack{i=1,\dots,n\\ j=1,\dots,r}}, \ \mathbf{R} = [\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)]_{i,j=1}^{n}.$$

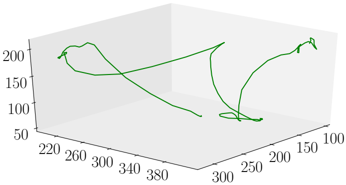
BIC

Bayesian Information Criteria is a trade-off between prediction quality and the number of selected features $\|\mathbf{a}\|_0$:

$$\text{BIC} = m \ln\left(\text{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathrm{a})\right) + \|\mathbf{a}\|_0 \cdot \log m.$$

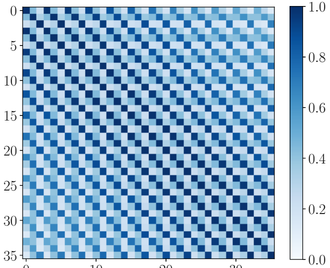
Computational experiment





$$\mathbf{X} \in \mathbb{R}^{m\times(32\cdot 27)}; \quad \mathbf{Y} \in \mathbb{R}^{m\times 3k}.$$

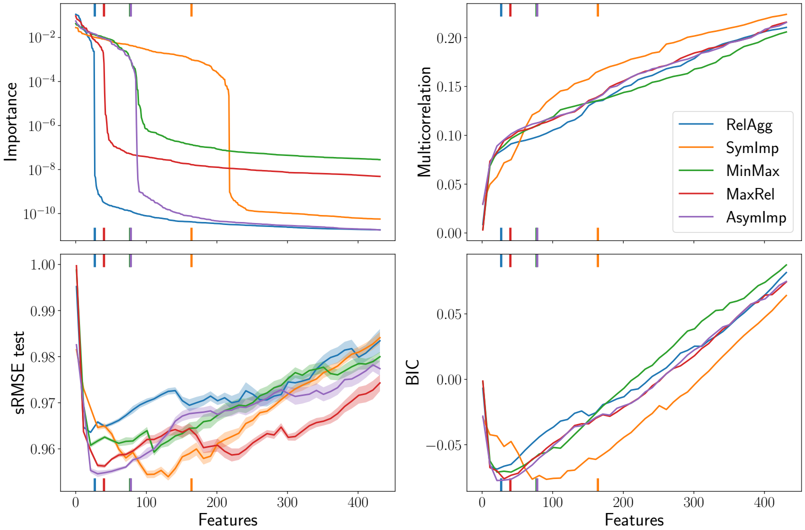
$$\mathbf{Y} = \begin{pmatrix} x_1 & y_1 & z_1 & \dots & x_k & y_k & z_k \\ x_2 & y_2 & z_2 & \dots & x_{k+1} & y_{k+1} & z_{k+1} \\ & \dots & & \dots & & \dots & \\ x_m & y_m & z_m & \dots & x_{m+k} & y_{m+k} & z_{m+k} \end{pmatrix}$$



Y correlation matrix

http://neurotycho.org

Quality criteria evaluation



Stability of selected feature subsets

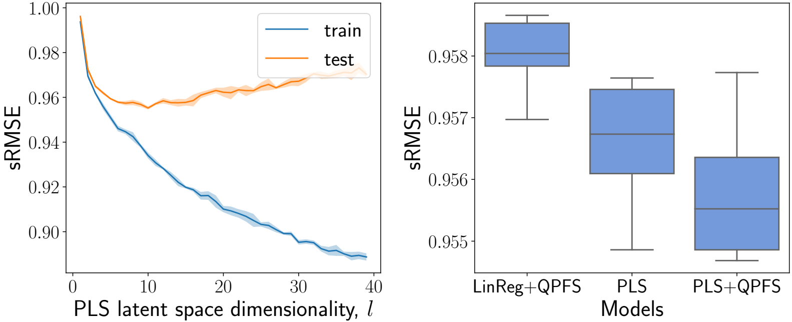
Experiment design

- generate bootstrap data
$$(\mathbf{X}, \mathbf{Y}) \to \{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\};$$
- solve feature selection problem
$$\{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\} \to \{\mathbf{z}_1, \dots, \mathbf{z}_s\};$$
- calculate statistics
$$\{\mathbf{z}_1, \dots, \mathbf{z}_s\} \to \{\text{RMSE}, \|\mathbf{a}\|_0, \text{Spearman } \rho, \ell_2 \text{ dist}\}.$$

	sRMSE	$\|\mathbf{a}\|_0$	Spearman ρ	ℓ_2 dist
RelAgg	0.965 $\pm$ 0.002	26.8 $\pm$ 3.8	0.915 $\pm$ 0.016	0.145 $\pm$ 0.018
SymImp	0.961 $\pm$ 0.001	224.4 $\pm$ 9.0	0.910 $\pm$ 0.017	0.025 $\pm$ 0.002
MinMax	0.961 $\pm$ 0.002	101.0 $\pm$ 2.1	0.932 $\pm$ 0.009	0.059 $\pm$ 0.004
MaxRel	0.958 $\pm$ 0.003	41.2 $\pm$ 5.2	0.862 $\pm$ 0.027	0.178 $\pm$ 0.010
AsymImp	0.955 $\pm$ 0.001	85.8 $\pm$ 10.2	0.926 $\pm$ 0.011	0.078 $\pm$ 0.007

Design of experiment

To compare feature selection and dimensionality reduction for linear regression and PLS regression models.



Results

- The problem of ECoG signal decoding in high dimensional spaces is investigated.
- Dimensionality reduction technique with space structure analysis is investigated.
- Feature selection methods which take into accout structure of both input and target spaces are proposed.
- The combination of feature selection and dimensionality reduction is proposed.
- Proposed feature selection algorithms give the stable and adequate solutions.

Conclusion

Publications

- Isachenko R., Strijov V. Metric learning for time series multiclass classification *Informatics and Applications*, 10(2), 2016.
- Isachenko R. et al. Feature Generation for Physical Activity Classification. *Artificial Intellegence and Decision Making*, 2018, accepted to the journal.
- Isachenko R., Strijov V. Quadratic programming optimization for Newton method. *Lobachevskii Journal of Mathematics*, 2018, accepted to the journal.
- Isachenko R., Strijov V. Dimensionality reduction for multivariate ECoG-based data. *Chemometrics*, 2018, ready for submission.

Conferences

- Lomonosov, 2016, Moscow. Metric learning in multiclass time series classification.
- Intelligent Data Processing Conference, 2016, Barcelona. Multimodel forecasting multiscale time series in internet of things.
- MMRO, 2017, Taganrog. Local models for classification of complex structured objects.