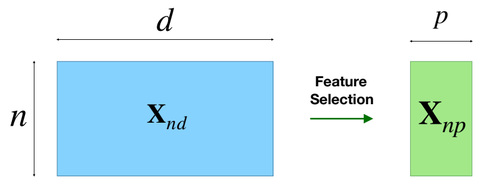
Kernel Latent regularization for feature selection in tumor classification

Martin Palazzo[123], Patricio Yankilevich[1] , Pierre Beauseroy[2]

[1]Biomedicine Research Institute of Buenos Aires - CONICET - Partner Institute of the Max Planck Society
[2] Université de Technologie de Troyes, France
[3] Universidad Tecnológica Nacional Buenos Aires

Feature Selection in tumor profiles

Given a gene signature tumor profiles can be classified by low and high survival rate. Tumors are characterized by more than d = 20.000 feature genes and cancer datasets use to have a low sample n size where n << d . To avoid the curse of dimensionality and to guide biomarker discovery feature selection methods are used and a subset of p genes are selected.



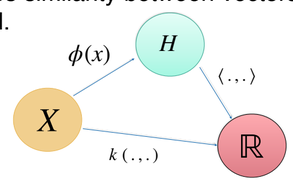
Instead of learning a selection function f(x) from a supervised approach y = f(x) which may be affected by overfitting we propose to learn a selection function from the labels and from the latent structure **z** of the training data simultaneously via kernel methods. The objective is to regularize the selection and to improve generalization in classification tasks.

$$y = \hat{f}(x) \rightarrow (z, y) = \hat{f}(x)$$

Kernel Methods

In this work Kernel functions are used to mix latent structure with supervised labels. A Kernel measures similarity between vectors in a reproducing kernel Hilbert space H.

$$k\left(x_i, x_j\right) = \left\langle \phi(x_i), \phi(x_j) \right\rangle_H$$



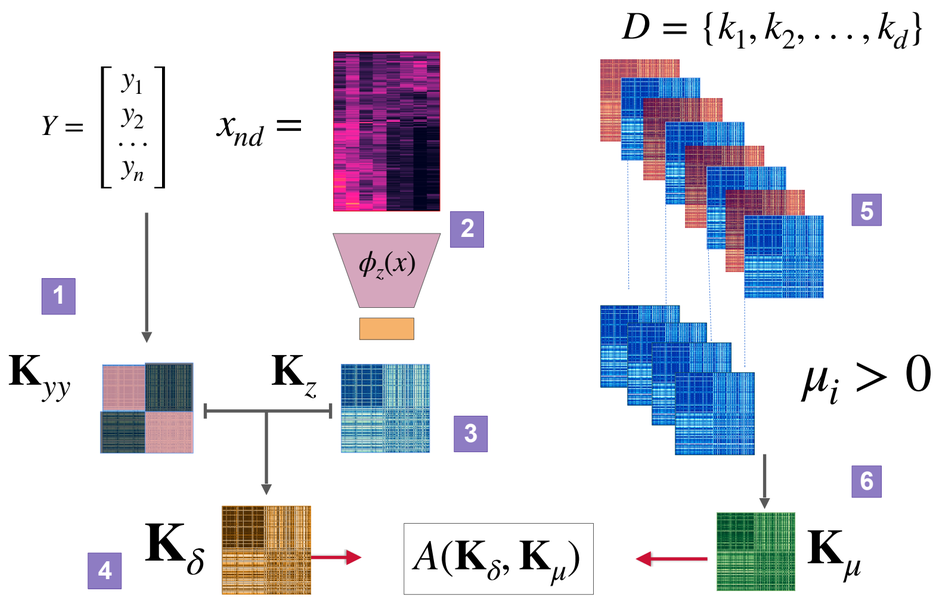
To learn from **z** and **y** kernel alignment is used. This score measures the similarity between a target kernel and a sample kernel. The higher the score the more similar both kernels.

$$A\left(K_1, K_2\right) = \frac{\left\langle K_1, K_2 \right\rangle_F}{\sqrt{\left\langle K_1, K_1 \right\rangle_F \left\langle K_2, K_2 \right\rangle_F}}$$

To select features multiple kernel learning is used. A linear combination of feature-wise kernels is done to obtain a new custom kernel that maximizes the alignment with the target kernel.

$$\boldsymbol{k_\mu}\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \sum_{i=1}^{n} \mu_i k_i\left(\boldsymbol{x}, \boldsymbol{x}'\right), \mu_i \geq 0$$

Proposed method: KLR-FS (arxiv.org/abs/2004.04866)



$$\mathbf{K}_\delta = (1-\delta)\mathbf{K}_z + \delta\mathbf{K}_y$$

$$\mathbf{K}_\mu = \sum_{i=1}^{n} \mu_i \mathbf{K}_i\left(\boldsymbol{x}, \boldsymbol{x}'\right)$$

$= x_{np}$

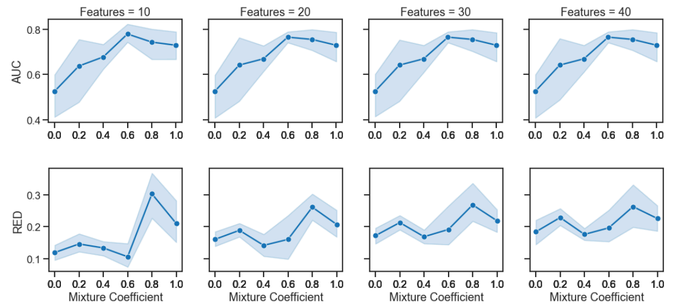
8

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{M} \alpha_i y_i k_\mu(\mathbf{x}, \mathbf{x}_i) + b$$

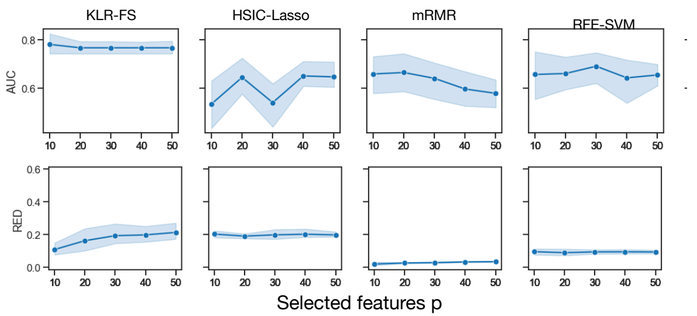
(1) A supervised Ky kernel is build with tumor labels. **(2)** A latent space from training data is obtained via a non-linear dimensionality reduction method (kPCA). **(3)** An unsupervised kernel Kz is built on the latent space. **(4)** Both kernels are mixed via a mixture coefficient Delta to obtain a target kernel that keeps supervised and unsupervised information. **(5)** A set of feature-wise kernels are built and linearly combined via multiple kernel learning (MKL). **(6)** The target kernel used in the MKL step is the delta kernel, then only the feature-wise kernels that approximate to the target kernel are kept and the rest discarded. **(7)** With the feature-weights higher than zero features are selected. **(8)** The kernel obtained from MKL is used in Support Vector Classification.

Results

The method is evaluated with the Breast Cancer patient survival data from the BRCA-US repository at the International Cancer Genome Consortium. Area Under the ROC Curve is used to evaluate tumor classification. The Redundancy score (RED) is used to evaluate how redundant are the selected features, the lower the RED the better the selection. The train-test split is done five times randomly and selection and classification results are averaged.



KLR-FS is evaluated for different values of selected features and different mixtures (delta parameter) of supervised and unsupervised target kernels. At delta = 0.6 there is a peak of AUC and simultaneously a low redundancy, suggesting that learning partially from latent structure improves the feature selection and the classification on test samples.



KLR-FS with delta=0.6 is compared with HSIC-Lasso, mRMR and RFE-SVM in support vector classification. The highest AUC for different values of p is achieved by the proposed method. Despite the fact that the RED of the selected features by KLR-FS is not the smallest, it is at similar levels as the benchmark methods.

Conclusions

Learning from both latent structure and tumor labels improves the classification performance and serves as regularization.