A kernel hat matrix based rejection criterion for outlier removal and data classification(F. Dufrenois)

Université du Littoral Côte d'Opale

Problematic and assumptions

- Outliers are omnipresent : economic science, astronomical science,… computer vision, etc..
- Statistics of the hat matrix used for outlier identification. We propose a nonlinear extension of the standard hat matrix : The **kernel hat matrix**
- A model selection criterion is derived for **dominant data subset extraction**
- Performances of the proposed approaches are studied on simulated and real data sets.

From the hat matrix to the kernel hat matrix

1 Hat matrix

Let $Z=[X{:}\mathbf{y}]=[\mathbf{z}_1, \mathbf{z}_2, \ldots \mathbf{z}_n]^T$ is the $n\times(d+1)$ matrix that combines both the matrix of predictors X and the response vectors $\mathbf{y}$. The augmented hat matrix is: $\mathcal{H}=Z\left(Z^TZ\right)^{-1}Z^T$

2 « Kernelized » Hat matrix

Consider a nonlinear transformation $\phi: \mathbb{R}^{d+1}\rightarrow\mathbb{R}^p$, $\mathbf{z}\rightarrow\phi(\mathbf{z})$ and the $n\times p$ matrix $\Phi=(\phi(\mathbf{z}_1), \phi(\mathbf{z}_2),\ldots,\phi(\mathbf{z}_n))^T$

$$\underbrace{H_\Phi=\Phi\,(\Phi^T\Phi)^{-1}\Phi^T}_{p\times p\ (\text{high dimension})} \xrightarrow{\lim_{\gamma\to 0}\left(Z^TZ+\gamma I\right)^{-1}Z^T=\lim_{\gamma\to 0}Z^T\left(ZZ^T+\gamma I\right)^{-1}} \underbrace{H_\Phi=\Phi\Phi^T\,(\Phi\Phi^T+\gamma I)^{-1}}_{n\times n\ (\text{low dimension})}$$

Regularization parameter

$K=\Phi.\Phi^T$

$$H_K=K\,(K+\gamma I)^{-1}$$

Eigenvalue decomposition :
$V=(V_1,V_2,..V_n)$ matrix of eigenvectors of K
$\Lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ matrix of eigenvalues of K

$$H_K=V\Omega V^T \qquad \Omega=\mathrm{diag}\left(\frac{\lambda_1}{\lambda_1+\gamma},\frac{\lambda_2}{\lambda_2+\gamma},\ldots,\frac{\lambda_n}{\lambda_n+\gamma}\right)$$

$$\mathrm{Trace}(H_K)$$

Effective degree of freedom

3 Gaussian kernel hat matrix and properties

$$K_\sigma(x, x_i)=e^{\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right)} \Rightarrow H_K(\sigma,\gamma)$$

- **How does H_K evolve with respect to (σ,γ) ???** We show that a pair (σ,γ) exists which best separates a dominant subpopluation from outliers !!! Look at this example of three unidimensional data points

A dominant subpopulation of two points (Z1 and Z2) and one outlier (Z3)

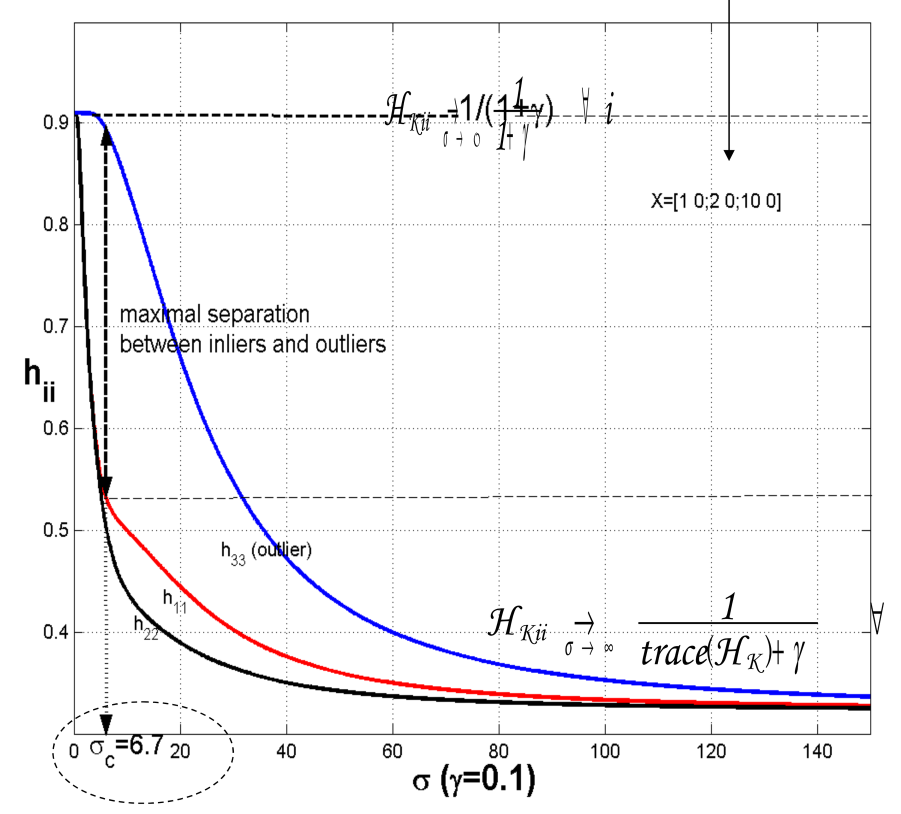
The evolution of the corresponding h_{ii} with respect to σ shows a maximal separation between h33 and the pair (h11, h22) for a critical value of σ (σ_c in the picture on the right) ! ! !

An analytical expression of σ_c can be obtained

$$\sigma_c=\frac{\left\|z_1-z_2\right\|^2}{2}\,\frac{\alpha^2-1}{\ln\alpha}$$

with

$$\left(\alpha^2=\frac{\left\|z_1-z_2\right\|^2}{\left\|z_1-z_3\right\|^2}\right)$$



4 Proposed model selection for dominant data set extraction:

- ***Goal*** : optimizing the separation beetwen the hii's distribution of the dominant data subset from the one of the outliers.

- ***Decision threshold***: $H_{Kii} \lessgtr trace(H_K)/n \quad 1<i<n$

This threshold gives an initial partition of the hii's distribution into two distributions :
1) The « Dominant »distribution $H^{(D)}$ of mean $\mu^{(D)}$ and variance $\sigma^{(D)}$.
2) The « Outlier « distribution $H^{(O)}$ of mean $\mu^{(O)}$ and variance $\sigma^{(O)}$.

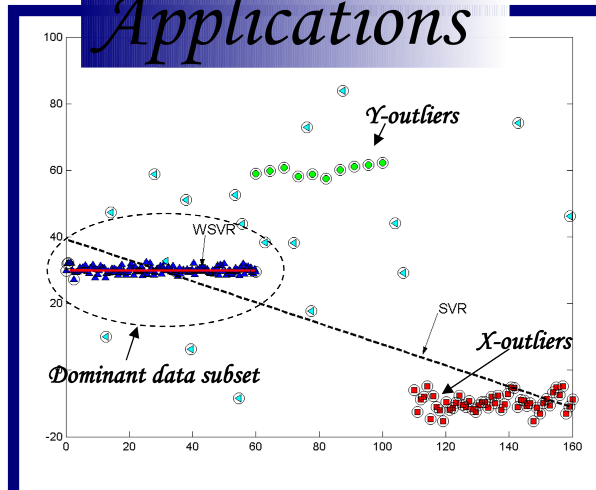
- ***Proposed model selection:*** We consider the couple (σ^*, γ^*) which maximizes the ratio between-subset variance (i.e separation) and within-subset variance (i.e overlap) :

$$\delta(\sigma,\gamma)=\frac{\left(\mu^{(D)}(\sigma,\gamma)-\mu^{(O)}(\sigma,\gamma)\right)^2}{\sigma^{(D)}(\sigma,\gamma)+\sigma^{(O)}(\sigma,\gamma)}$$

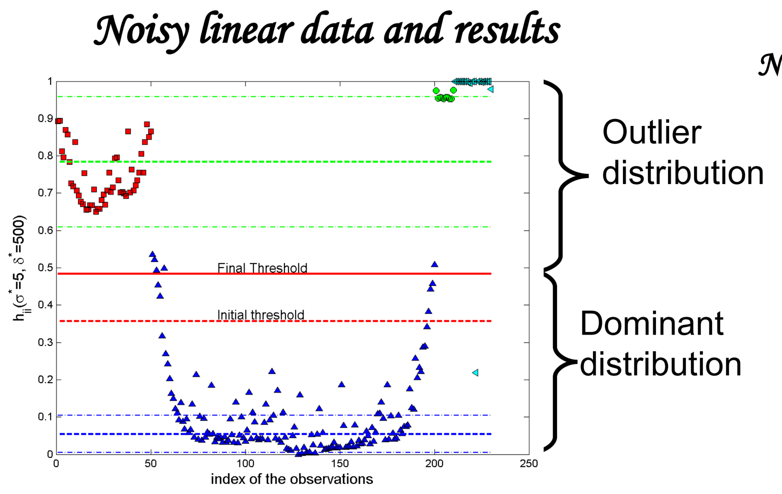
Linear Fisher's discriminant

5 Conclusion : The proposed model selection is used as learning stage for data classification before regression task: application to support vector regression : the proposed approach is called : Weighted SVR (WSVR)

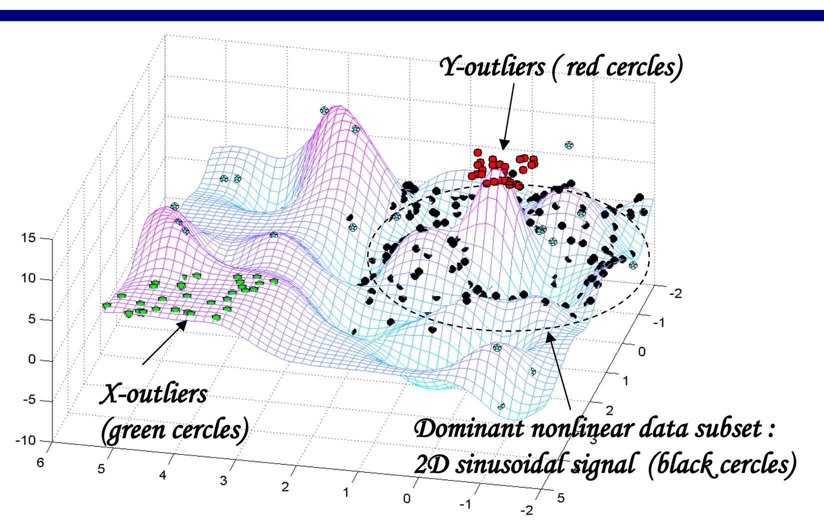
Applications



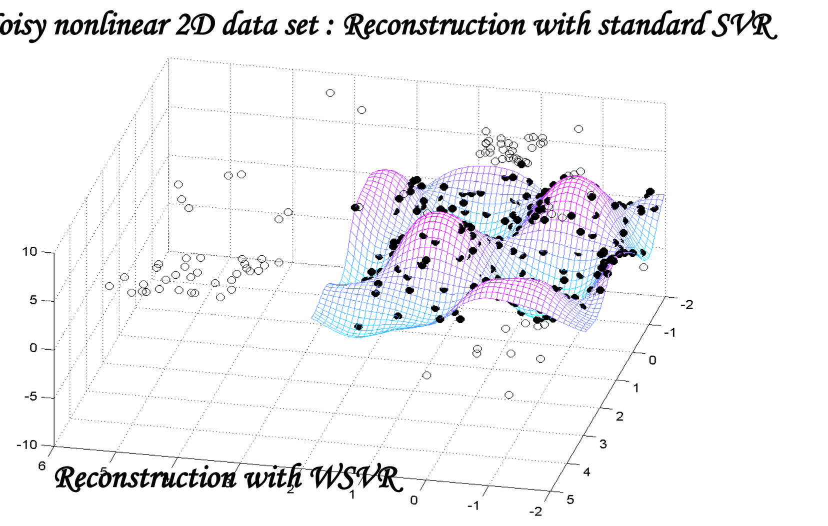
Noisy linear data and results



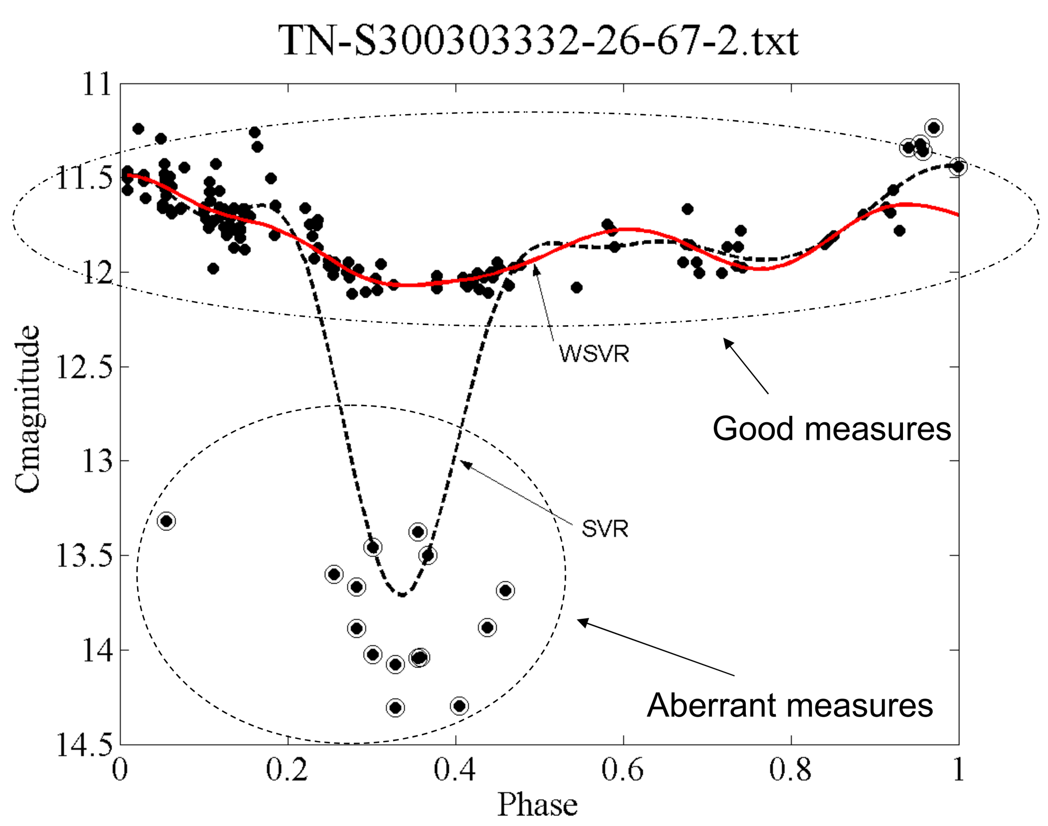
Hii's distribution and classification



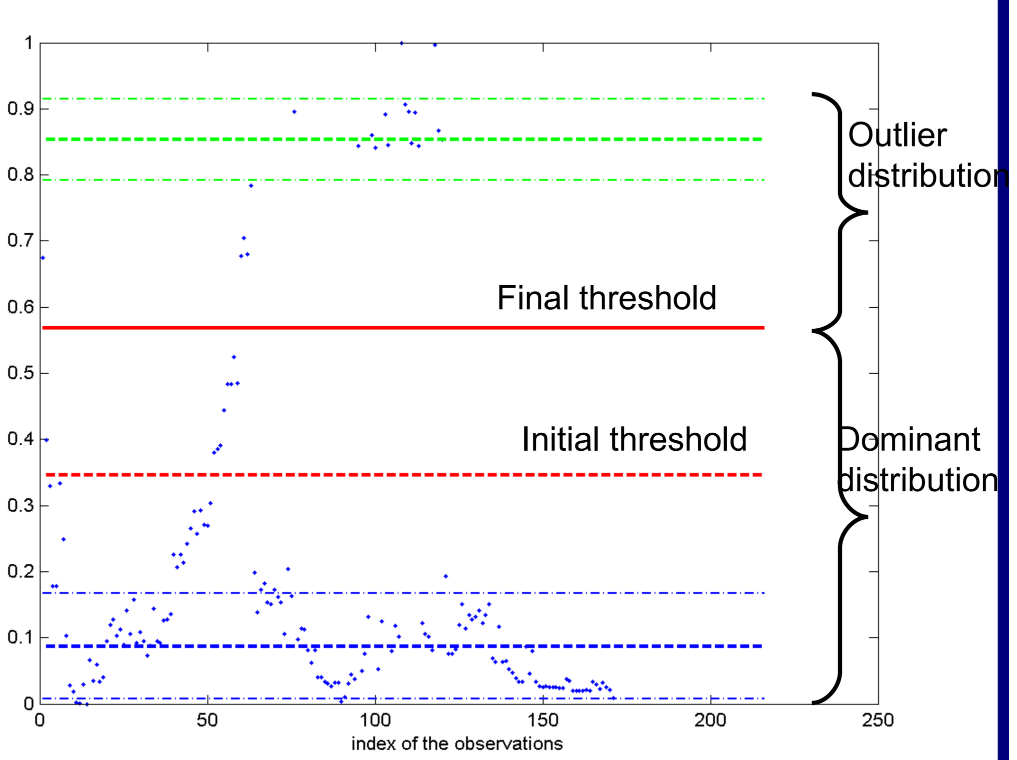
Noisy nonlinear 2D data set : Reconstruction with standard SVR



Reconstruction with WSVR



Real data set (TAROT data base) : Stars data with aberrant measures. Standard SVR is attracted by outliers while WSVR gives a correct reconstruction



Hii's distribution and classification of the star data set

Conclusion

- A new robust learning strategy for outlier detection has been proposed.
- The Gaussian Kernel hat matrix presents discriminative properties under the condition to choose appropriate values for kernel parameters.

Futur work

- Automatic selection of kernel parameters
- Spectral analysis of the kernel hat matrix for multiple modality extraction
- Application to computer vision problems : optical flow estimation, segmentation