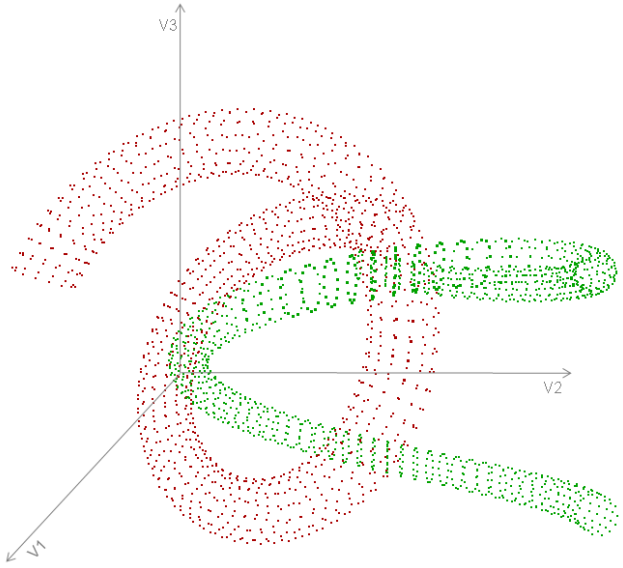
PerTurbo: a new classification algorithm based on the spectrum perturbations of the Laplace-Beltrami operator

Nicolas Courty, **Thomas Burger** and Johann Laurent

Université de Bretagne Sud

Problem and motivations

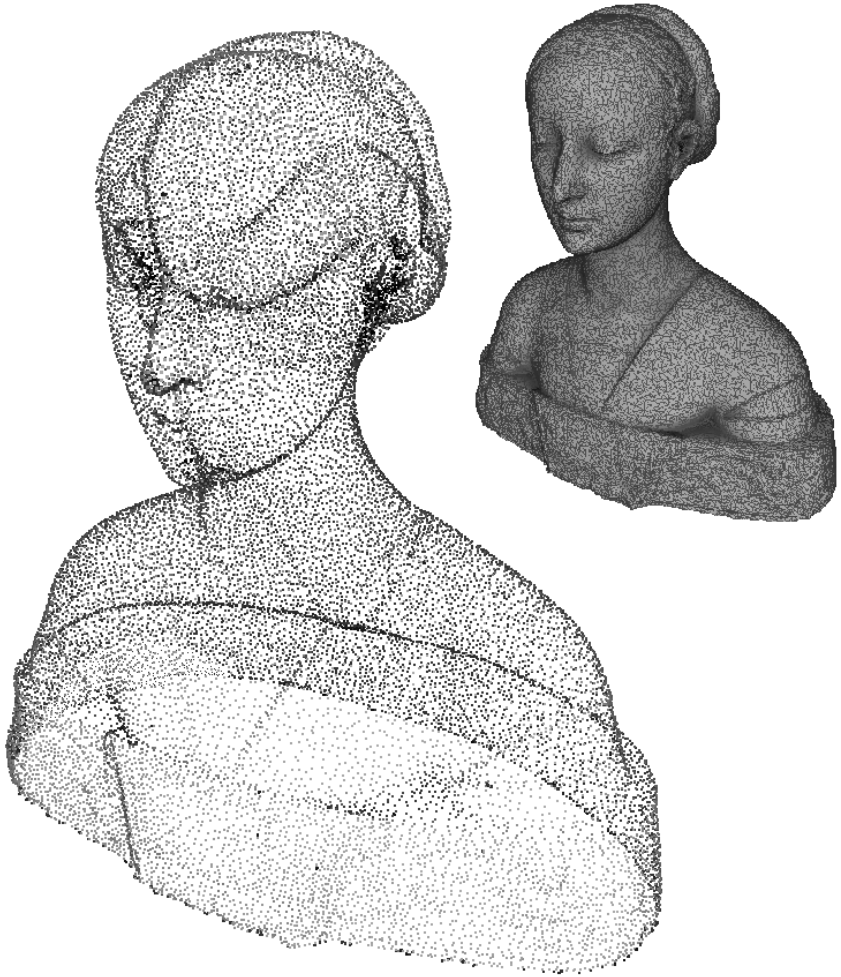
1. Supervised classification problem
2. A geometric point of view on the manifold processing
3. In 3 dimension problems, strong parallel with computer graphics



Outline

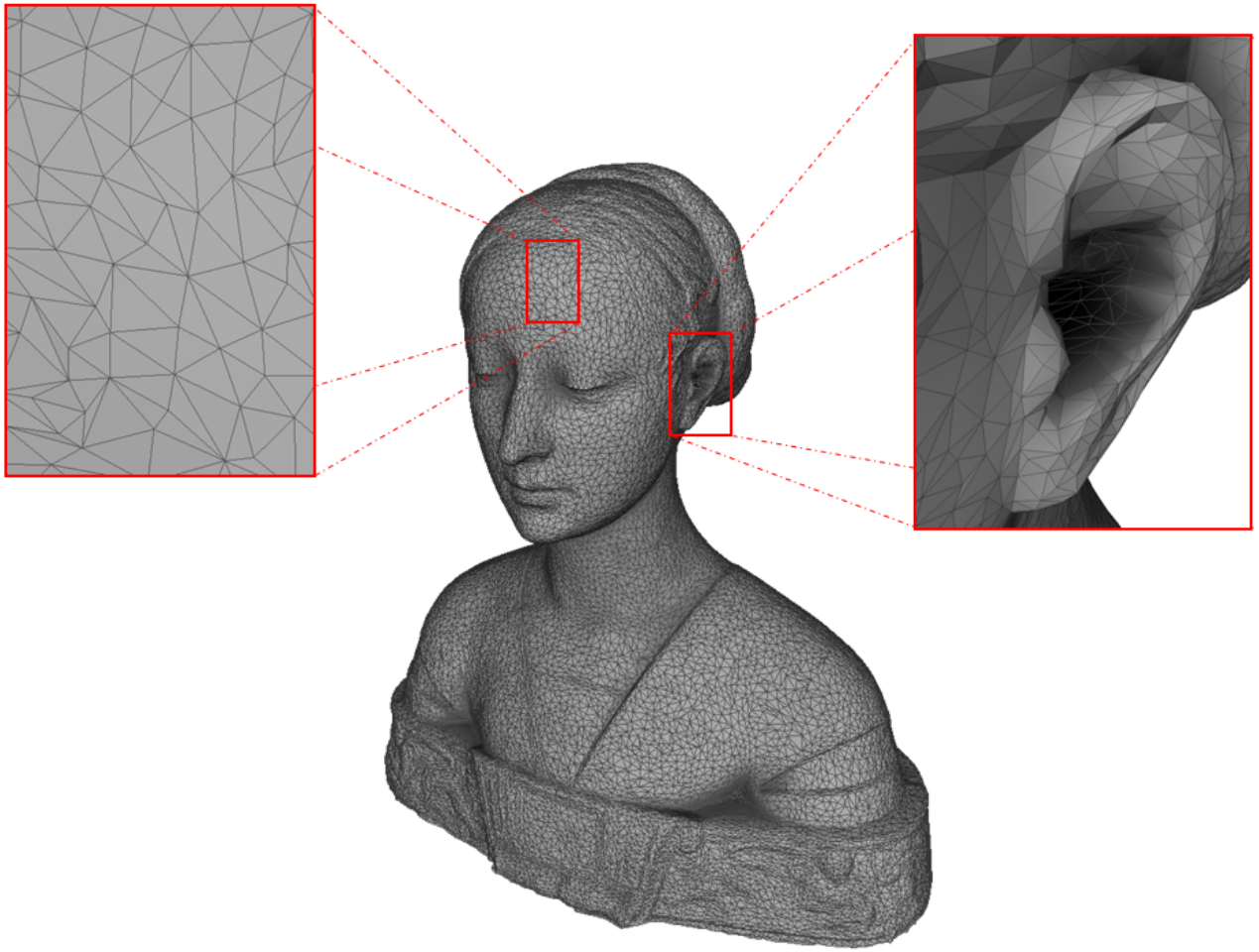
Outline

Sampling point clouds in computer graphics



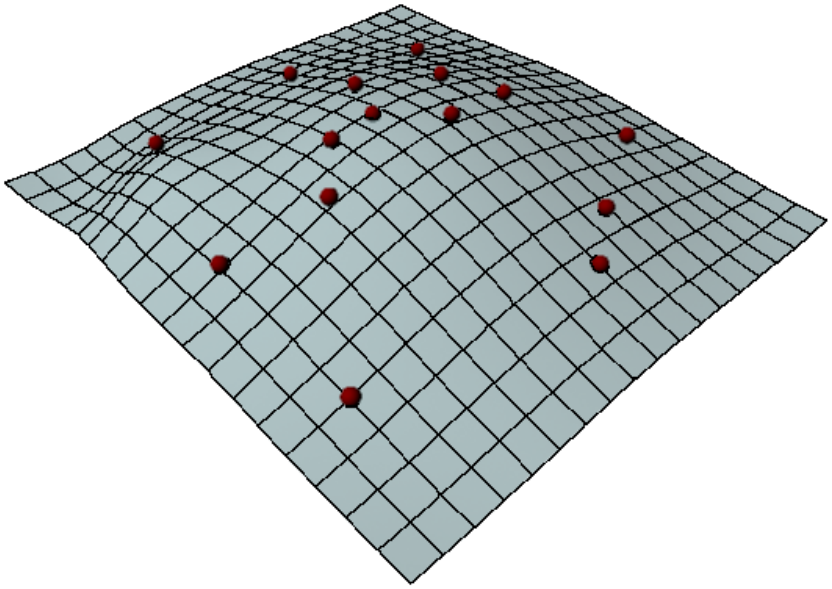
1. Too many points required to sample real life objects
2. The reduction of the number of samples leads to compact storing
3. Only the informative enough samples are kept

Spectral sampling



Need for metric to quantify the modification of the definition of the object surface induced by a new sample [Öztireli, Alexa & Gross, 2010].

Riemanian manifolds and the Laplace-Beltrami Operator



1. The surface is assumed to be a Riemannian manifold $\mathscr{M}$ (i.e. differentiable everywhere).
2. The Laplace-Beltrami operator $\triangle(.)$ is a generalization of the Laplace operator for Riemannian manifolds.
3. The spectrum of $\triangle(\mathscr{M})$ completely defines $\mathscr{M}$ up to an isometry.

Approximating the Laplace-Beltrami operator

1. The surface being unknown, the computation of $\triangle(\mathscr{M})$ is impossible.

2. It can be approximated with the Gram matrix $\mathbf{K}$ of the samples $\{x_1, \ldots, x_i, \ldots x_N\}$, fitted with a Gaussian dissimilarity measure:

$$\mathbf{K}_{ij} = k(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) \tag{1}$$

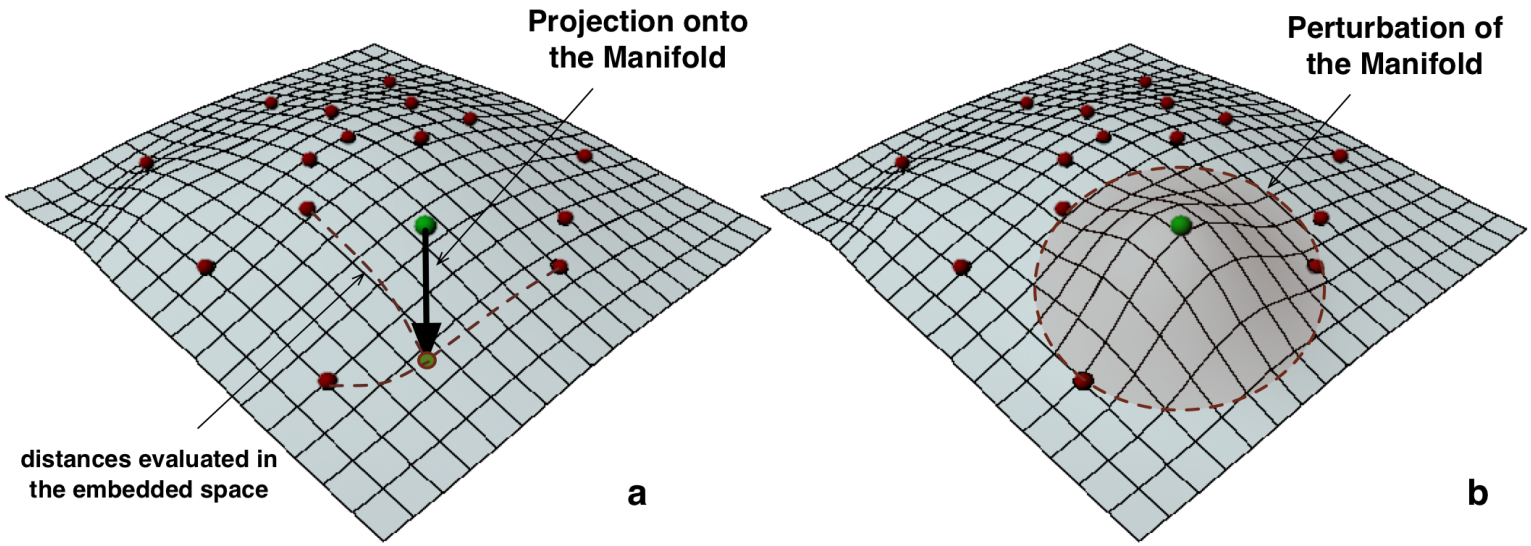
3. The spectrum of $\mathbf{K}$ is used to characterize the real object surface [Coifman & Lafon, 2006] and its pertubation is used to evaluate the interest of a sample [Öztireli, Alexa & Gross, 2010].

Outline

Idea #1 : Comparing a class to a surface

The pertubation of **K**:

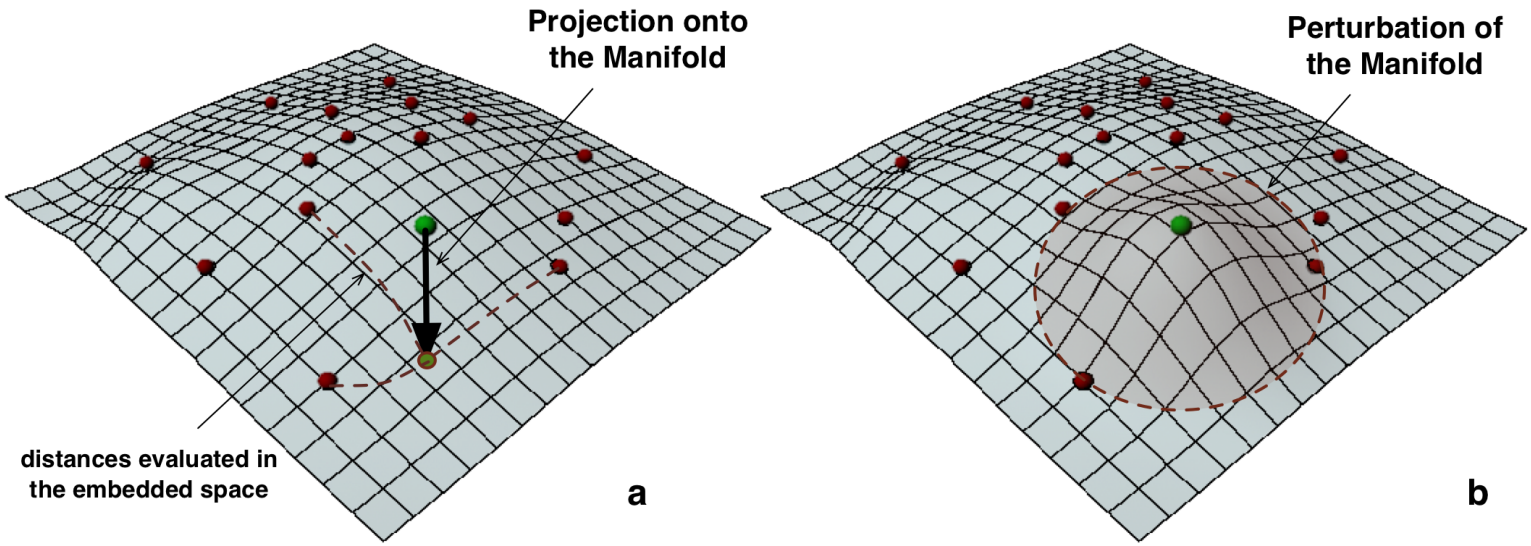
1. In computer graphics, it quantifies the interest of a sample.



Idea #1 : Comparing a class to a surface

The pertubation of **K**:

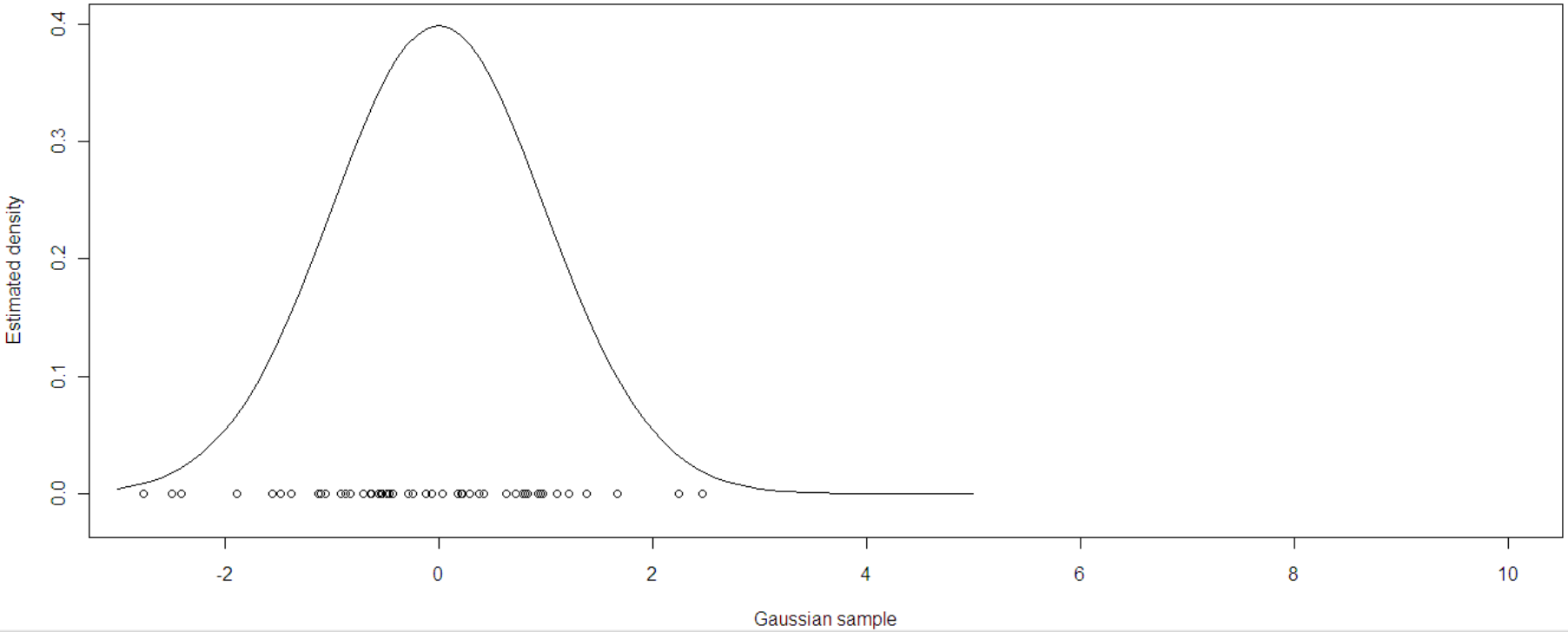
1. In computer graphics, it quantifies the interest of a sample.
2. In statistics ?



Idea #1 : Comparing a class to a surface

The pertubation of **K**:

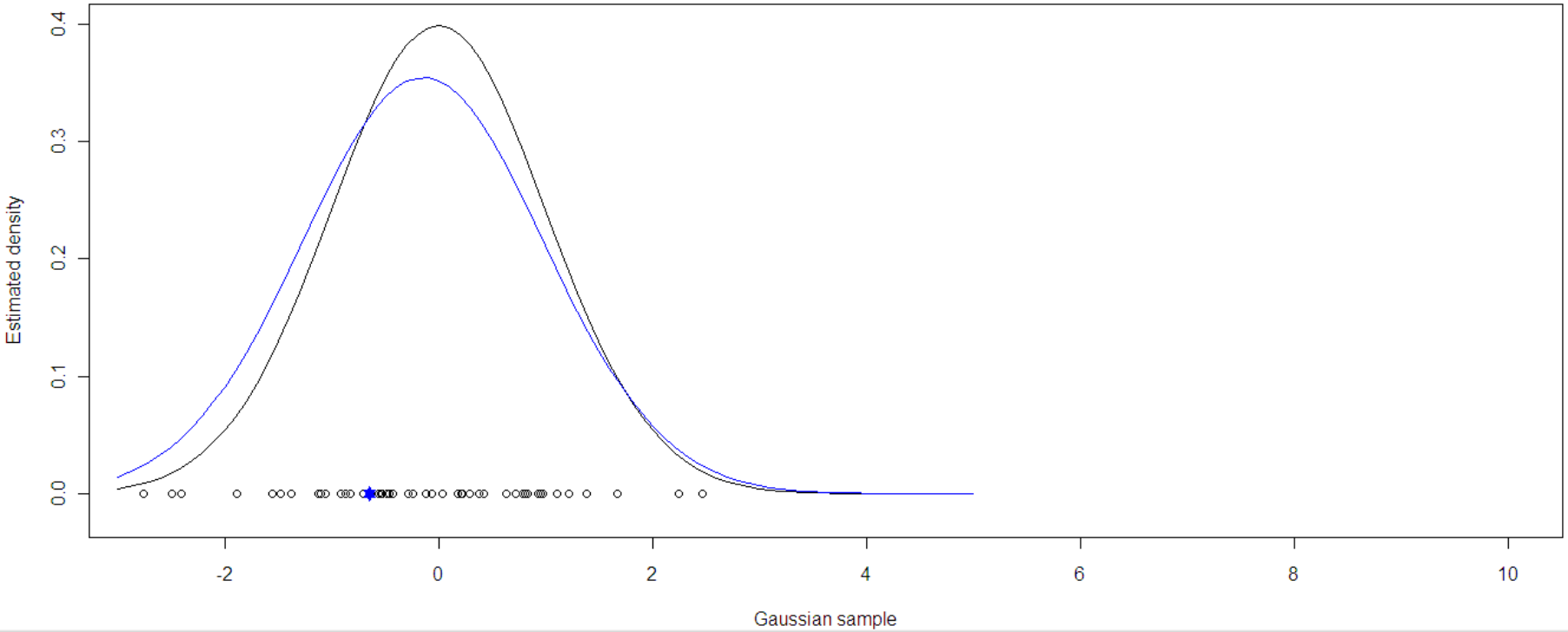
1. In computer graphics, it quantifies the interest of a sample.
2. In statistics ?
   1. Only an outlier modifies the distribution of its class



Idea #1 : Comparing a class to a surface

The pertubation of **K**:

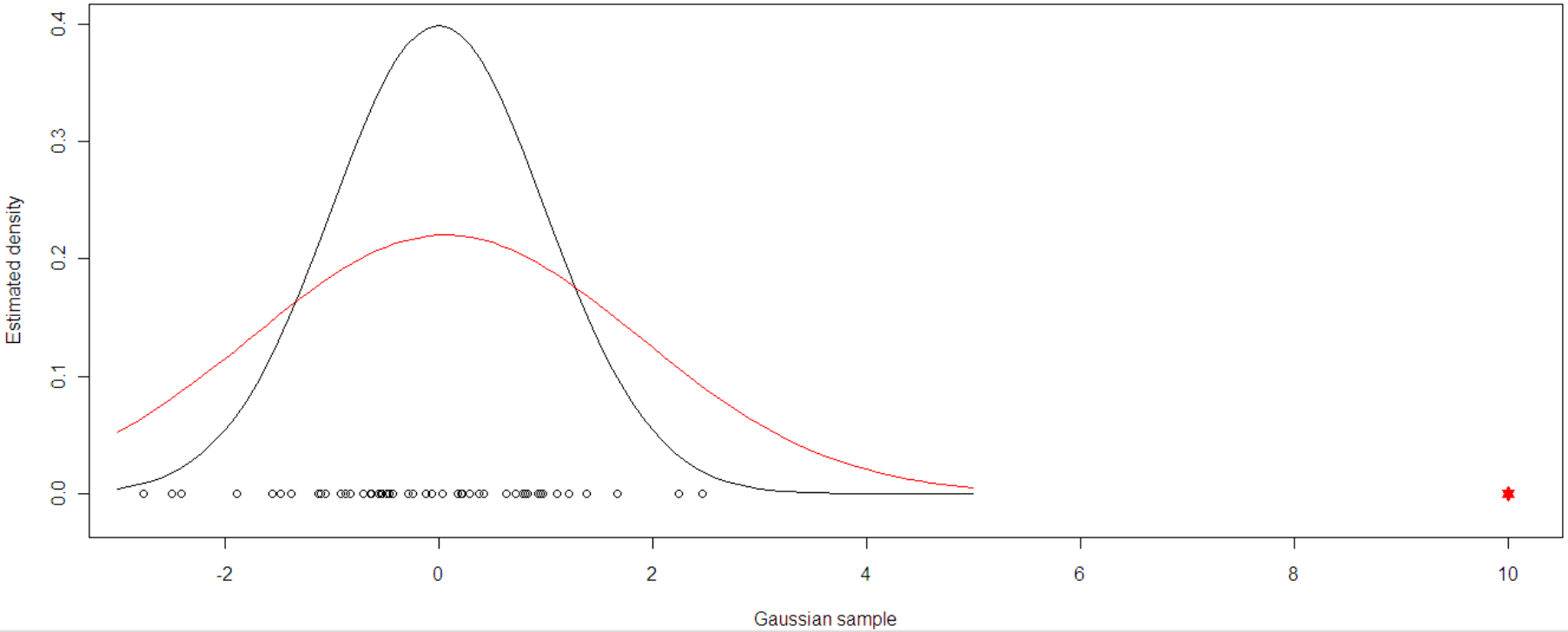
1. In computer graphics, it quantifies the interest of a sample.
2. In statistics ?
   1. Only an outlier modifies the distribution of its class



Idea #1 : Comparing a class to a surface

The pertubation of **K**:

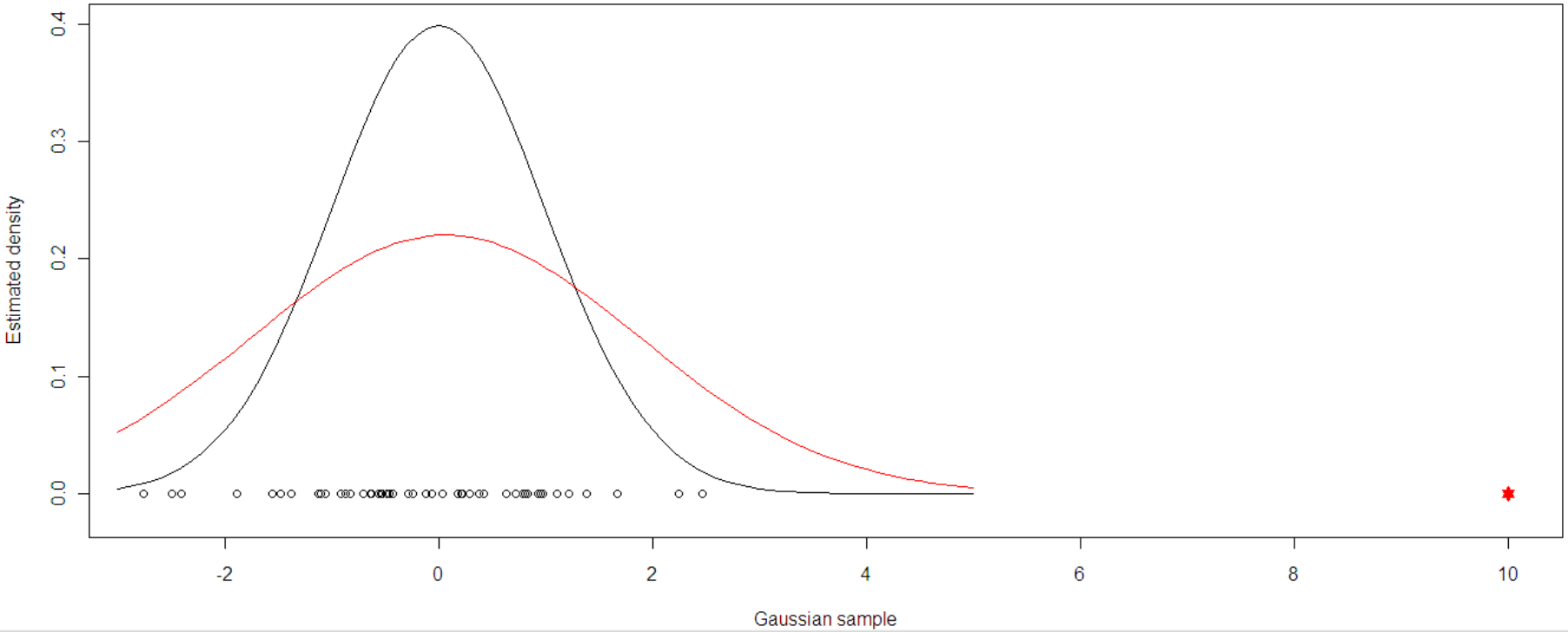
1. In computer graphics, it quantifies the interest of a sample.
2. In statistics ?
   1. Only an outlier modifies the distribution of its class



Idea #1 : Comparing a class to a surface

The pertubation of **K**:

1. In computer graphics, it quantifies the interest of a sample.
2. In statistics ?
   1. Only an outlier modifies the distribution of its class
   2. An interesting clue of the membership of a sample to a class !



Idea #2: Class-wise manifold learning

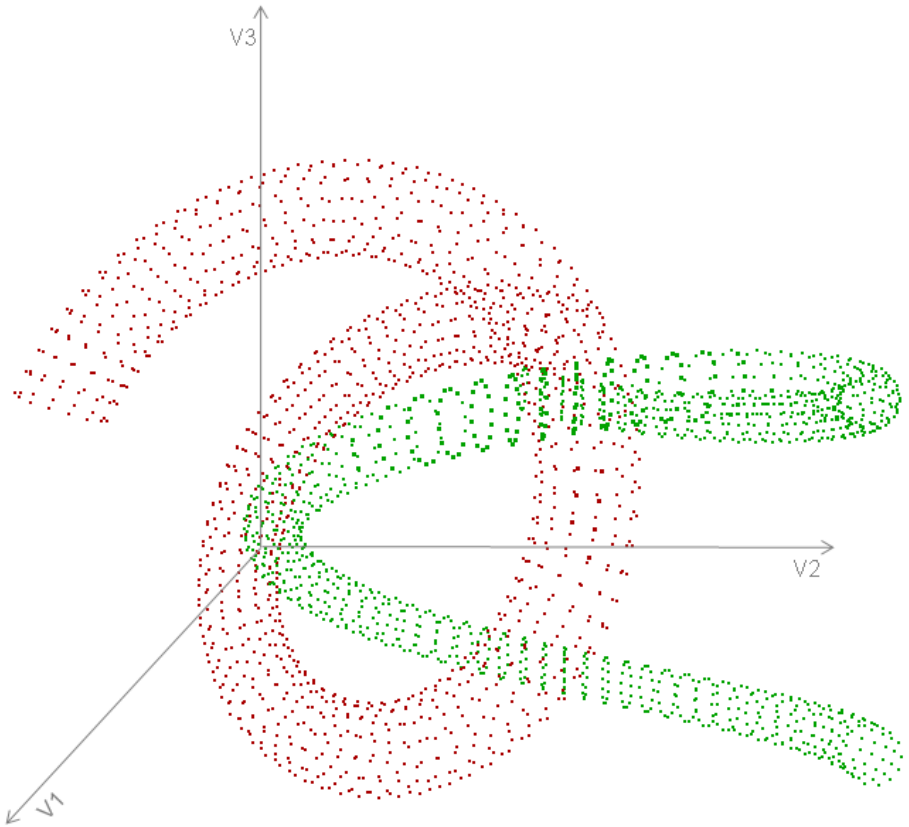


Figure: Fictive example of a 2-class problem where each class is embed in a dedicated manifold.

1. For each class ℓ, a dedicated manifold $\mathscr{M}_\ell$ is considered
2. All the manifold are learned independently
3. Each class ℓ is represented by a particular matrix $\mathbf{K}_\ell$

Remarks and notations

Remarks:

1. $\mathbf{K}_\ell$ is also the classical Gram matrix in the Gaussian RKHS
2. The proxy for Laplace-Beltrami operator $\equiv$ a kernel trick

Notations:

$\mathcal{X}$	The input space (generated by the problem variables)
$\mathcal{Z}$	The feature space associated to the Gaussian kernel
$\phi()$	the mapping from $\mathcal{X}$ onto $\mathcal{Z}$
$\mathcal{T}_\ell$	The set of training examples for $\mathscr{M}_\ell$
$\mathbf{K}_\ell$	The Gram matrix of $\phi(\mathcal{T}_\ell)$ in $\mathcal{Z}$
$\tilde{x}$	A test sample in $\mathcal{X}$

The perturbation measure in the Gaussian RKHS (1)

Let us project $\phi(\tilde{x})$ on the manifold $\mathscr{M}_\ell$ in $\mathcal{Z}$, i.e. $< \phi(\mathcal{T}_\ell) >$:

$$\phi(\tilde{x}) = \underbrace{r_\ell(\tilde{x})}_{\text{coplanar to } <\phi(\mathcal{T}_\ell)>} + \underbrace{o_\ell(\tilde{x})}_{\text{orthogonal to } <\phi(\mathcal{T}_\ell)>} \tag{2}$$

The perturbation measure in the Gaussian RKHS (1)

Let us project $\phi(\tilde{x})$ on the manifold $\mathscr{M}_\ell$ in $\mathcal{Z}$, i.e. $< \phi(\mathcal{T}_\ell) >$:

$$\phi(\tilde{x}) = \underbrace{r_\ell(\tilde{x})}_{\text{coplanar to } <\phi(\mathcal{T}_\ell)>} + \underbrace{o_\ell(\tilde{x})}_{\text{orthogonal to } <\phi(\mathcal{T}_\ell)>} \tag{2}$$

From [Öztireli, Alexa & Gross, 2010], the perturbation of $\mathbf{K}_\ell$ by $\tilde{x}$ comes from $o_\ell(\tilde{x})$. Thus, the perturbation measure reads:

$$\tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) = \frac{||o_\ell(\tilde{x})||^2}{||\phi(\tilde{x})||^2} = 1 - \frac{||r_\ell(\tilde{x})||^2}{||\phi(\tilde{x})||^2} = 1 - ||r_\ell(\tilde{\mathbf{x}})||^2 \tag{3}$$

The perturbation measure in the Gaussian RKHS (1)

Let us project $\phi(\tilde{x})$ on the manifold $\mathscr{M}_\ell$ in $\mathcal{Z}$, i.e. $< \phi(\mathcal{T}_\ell) >$:

$$\phi(\tilde{x}) = \underbrace{r_\ell(\tilde{x})}_{\text{coplanar to } <\phi(\mathcal{T}_\ell)>} + \underbrace{o_\ell(\tilde{x})}_{\text{orthogonal to } <\phi(\mathcal{T}_\ell)>} \tag{2}$$

From [Öztireli, Alexa & Gross, 2010], the perturbation of $\mathbf{K}_\ell$ by $\tilde{x}$ comes from $o_\ell(\tilde{x})$. Thus, the perturbation measure reads:

$$\tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) = \frac{||o_\ell(\tilde{x})||^2}{||\phi(\tilde{x})||^2} = 1 - \frac{||r_\ell(\tilde{x})||^2}{||\phi(\tilde{x})||^2} = 1 - ||r_\ell(\tilde{\mathbf{x}})||^2 \tag{3}$$

If Φ_ℓ is the matrix whose columns are the elements of $\phi(\mathcal{T}_\ell)$, then, the projector over $< \phi(\mathcal{T}_\ell) >$ can be written as: $\Phi_\ell(\Phi_\ell^T \Phi_\ell)^{-1}\Phi_\ell^T$

The perturbation measure in the Gaussian RKHS (2)

$$||r_\ell(\tilde{\mathbf{x}})||^2 = ||\Phi_\ell(\Phi_\ell^T\Phi_\ell)^{-1}\Phi_\ell^T\phi(\tilde{\mathbf{x}})||^2 \tag{4}$$

The perturbation measure in the Gaussian RKHS (2)

$$
\begin{aligned}
||r_\ell(\tilde{\mathbf{x}})||^2 &= ||\Phi_\ell(\Phi_\ell^T\Phi_\ell)^{-1}\Phi_\ell^T\phi(\tilde{\mathbf{x}})||^2 \\
&\vdots \\
&= (\phi(\tilde{\mathbf{x}})^T\Phi_\ell)\ ((\Phi_\ell^T\Phi_\ell)^T)^{-1}\ (\Phi_\ell^T\phi(\tilde{\mathbf{x}})) \qquad (4)
\end{aligned}
$$

The perturbation measure in the Gaussian RKHS (2)

$$
\begin{aligned}
||r_\ell(\tilde{\mathbf{x}})||^2 &= ||\Phi_\ell(\Phi_\ell^T\Phi_\ell)^{-1}\Phi_\ell^T\phi(\tilde{\mathbf{x}})||^2 \\
&\vdots \\
&= (\phi(\tilde{\mathbf{x}})^T\Phi_\ell)\ \underbrace{((\Phi_\ell^T\Phi_\ell)^T)^{-1}}_{\mathbf{K}_\ell^{-1}}\ (\Phi_\ell^T\phi(\tilde{\mathbf{x}})) \qquad (4)
\end{aligned}
$$

The perturbation measure in the Gaussian RKHS (2)

$$
\begin{aligned}
||r_\ell(\tilde{\mathbf{x}})||^2 &= ||\Phi_\ell(\Phi_\ell^T\Phi_\ell)^{-1}\Phi_\ell^T\phi(\tilde{\mathbf{x}})||^2 \\
&\vdots \\
&= \underbrace{(\phi(\tilde{\mathbf{x}})^T\Phi_\ell)}_{\mathbf{k}_\ell^T(\tilde{\mathbf{x}})}\ \underbrace{((\Phi_\ell^T\Phi_\ell)^T)^{-1}}_{\mathbf{K}_\ell^{-1}}\ \underbrace{(\Phi_\ell^T\phi(\tilde{\mathbf{x}}))}_{\mathbf{k}_\ell(\tilde{\mathbf{x}})} \qquad (4)
\end{aligned}
$$

The perturbation measure in the Gaussian RKHS (2)

$$
\begin{aligned}
||r_\ell(\tilde{\mathbf{x}})||^2 &= ||\Phi_\ell(\Phi_\ell^T\Phi_\ell)^{-1}\Phi_\ell^T\phi(\tilde{\mathbf{x}})||^2 \\
&\vdots \\
&= \underbrace{(\phi(\tilde{\mathbf{x}})^T\Phi_\ell)}_{\mathbf{k}_\ell^T(\tilde{\mathbf{x}})}\ \underbrace{((\Phi_\ell^T\Phi_\ell)^T)^{-1}}_{\mathbf{K}_\ell^{-1}}\ \underbrace{(\Phi_\ell^T\phi(\tilde{\mathbf{x}}))}_{\mathbf{k}_\ell(\tilde{\mathbf{x}})} \qquad (4)
\end{aligned}
$$

Finally:

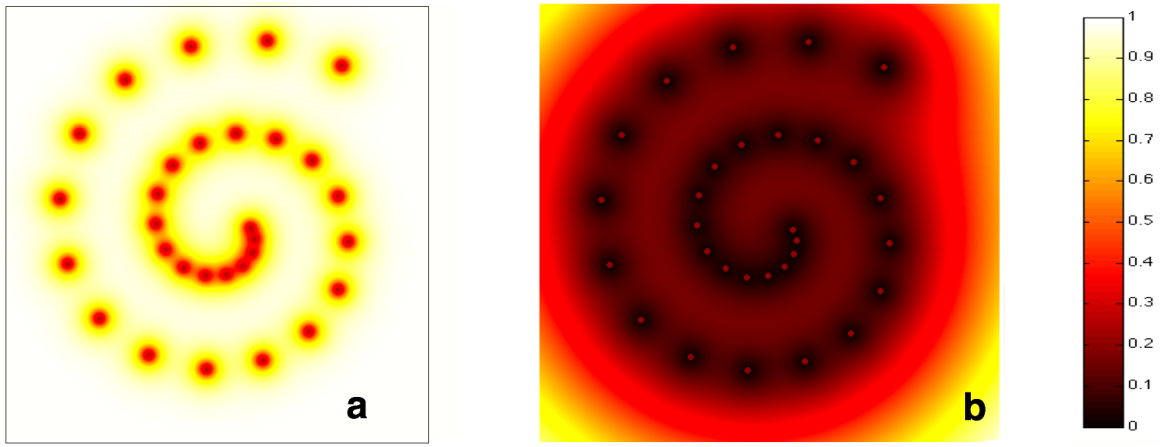
$$
\tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) = 1 - \mathbf{k}_\ell^T(\tilde{\mathbf{x}})\mathbf{K}_\ell^{-1}\mathbf{k}_\ell(\tilde{\mathbf{x}}) \qquad (5)
$$

The perturbation measure in the Gaussian RKHS (2)

$$
\begin{aligned}
||r_\ell(\tilde{\mathbf{x}})||^2 &= ||\Phi_\ell(\Phi_\ell^T\Phi_\ell)^{-1}\Phi_\ell^T\phi(\tilde{\mathbf{x}})||^2 \\
&\vdots \\
&= \underbrace{(\phi(\tilde{\mathbf{x}})^T\Phi_\ell)}_{\mathbf{k}_\ell^T(\tilde{\mathbf{x}})}\ \underbrace{((\Phi_\ell^T\Phi_\ell)^T)^{-1}}_{\mathbf{K}_\ell^{-1}}\ \underbrace{(\Phi_\ell^T\phi(\tilde{\mathbf{x}}))}_{\mathbf{k}_\ell(\tilde{\mathbf{x}})} \qquad (4)
\end{aligned}
$$

Finally:

$$
\tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) = 1 - \mathbf{k}_\ell^T(\tilde{\mathbf{x}})\mathbf{K}_\ell^{-1}\mathbf{k}_\ell(\tilde{\mathbf{x}}) \qquad (5)
$$



PerTurbo: A new classification algorithm

Training step:

$\forall$ class ℓ, $\mathbf{K}_\ell^{-1}$ is computed.

PerTurbo: A new classification algorithm

Training step:

$\forall$ class ℓ, $\mathbf{K}_\ell^{-1}$ is computed.

Testing step:

1. The dissimilarity of a new test sample $\tilde{\mathbf{x}}$ to each class ℓ is derived from the perturbation of $\mathscr{M}_\ell$ induced by $\tilde{\mathbf{x}}$:

$$\tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) = 1 - \mathbf{k}_\ell^T(\tilde{\mathbf{x}})\mathbf{K}_\ell^{-1}\mathbf{k}_\ell(\tilde{\mathbf{x}}) \tag{6}$$

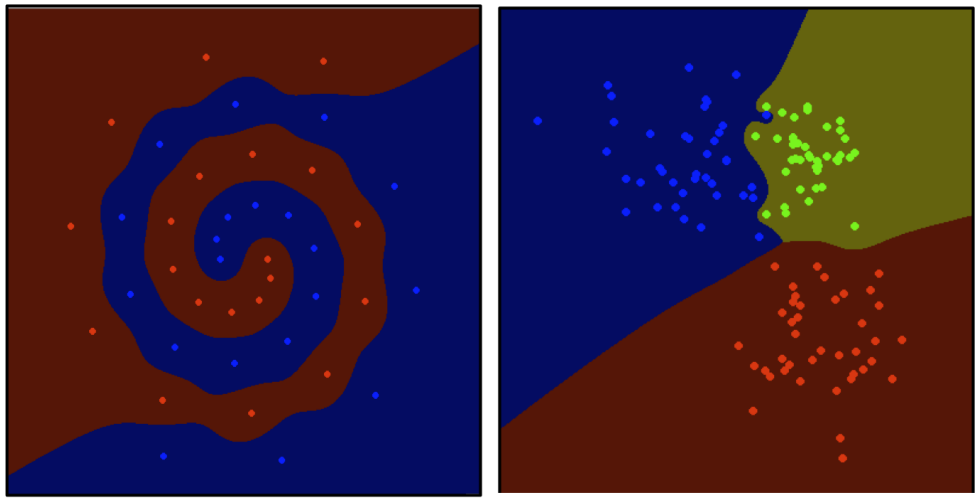
2. The sample $\tilde{\mathbf{x}}$ is associated to the class with the least induced perturbation, which reads as:

$$\arg\min_\ell \tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) \tag{7}$$

PerTurbo: A new classification algorithm

Training step:

$\forall$ class ℓ, $\mathbf{K}_\ell^{-1}$ is computed.



Testing step:

1. The dissimilarity of a new test sample $\tilde{\mathbf{x}}$ to each class ℓ is derived from the perturbation of $\mathscr{M}_\ell$ induced by $\tilde{\mathbf{x}}$:

$$\tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) = 1 - \mathbf{k}_\ell^T(\tilde{\mathbf{x}})\mathbf{K}_\ell^{-1}\mathbf{k}_\ell(\tilde{\mathbf{x}}) \tag{6}$$

2. The sample $\tilde{\mathbf{x}}$ is associated to the class with the least induced perturbation, which reads as:

$$\arg\min_\ell \tau(\tilde{\mathbf{x}}, \mathscr{M}_\ell) \tag{7}$$

Outline

Experimental setting and analysis of the results

Experimental setting:

1. Several tests conducted on simulated and real datasets [UCI Machine Learning Repository]
2. Comparison to several algorithms (among which SVMs)
3. SVMs are fully optimized with cross validation
4. Several versions of PerTurbo are tested (see article)

Experimental setting and analysis of the results

Experimental setting:

1. Several tests conducted on simulated and real datasets [UCI Machine Learning Repository]
2. Comparison to several algorithms (among which SVMs)
3. SVMs are fully optimized with cross validation
4. Several versions of PerTurbo are tested (see article)

Qualitative results of PerTurbo:

1. Performances similar to SVMs
2. Less efficient with missing values or binary variables
3. Depending on the problem, the best version is not the same

Datasets

Table: Description of simulated/UCI datasets

Datasets	#Training	#Tests	#Classes	#Variables	Comments
SimData-1	200	800	10	19	64 components
SimData-2	200	800	10	26	64 components
SimData-3	200	800	10	31	75 components
Ionosphere	71	280	2	34	
diabets	154	614	2	8	missing values
Blood-transfusion	150	598	2	4	
Ecoli	67	269	8	7	too small for CV
Glasses	43	171	6	9	
Wines	36	142	3	13	
Parkinsons	39	156	2	22	
Letter-reco	4000	16000	26	16	
Hill-valley1	606	606	2	100	50% unlabeled
Hill-valley2	606	606	2	100	50% unlabeled

Accuracy rates

Table: Comparison of the accuracy rates (mean and standard deviation, in percentages) with SVM (with optimized parameters and hyper-parameters)

Datasets	PerTurbo	SVM	Comparison
SimData-1	79.0 (1.8)	76.7 (1.3)	●
SimData-2	54.7 (2.5)	45.0 (1.8)	●
SimData-3	19.7 (1.2)	16.2 (1.3)	●
Ionosphere	92.1 (1.6)	92.5 (1.8)	●
diabets	72.6 (2.2)	74.0 (1.7)	●
Blood-transfusion	76.9 (1.0)	77.6 (1.5)	●
Ecoli	83.7 (2.5)	83.2 (2.2)	●
Glass	65.4 (2.9)	60.6 (4.4)	●
Wines	72.60 (1.3)	96.1 (1.7)	●
Parkinsons	83.8 (1.3)	85.0 (3.3)	●
Letter-reco	92.7 (0.2)	91.9 (0.3)	●
Hill-valley1	60.7	56.4	●
Hill-valley2	59.9	55.3	●

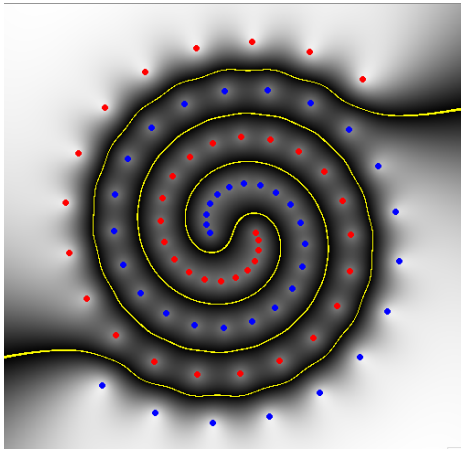
Outline

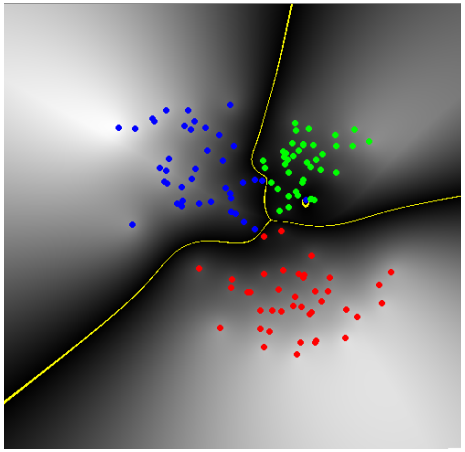
Active Learning (1)

It rather simple to find the borders of the classes:

$$B = \{\mathbf{x} \in \mathcal{X} \text{ s.t. } |\tau_{r(1)}(\mathbf{x}) - \tau_{r(2)}(\mathbf{x})| = 0\}$$

where $r(i)$ is the ith least perturbated classes by x.





Active Learning (1)

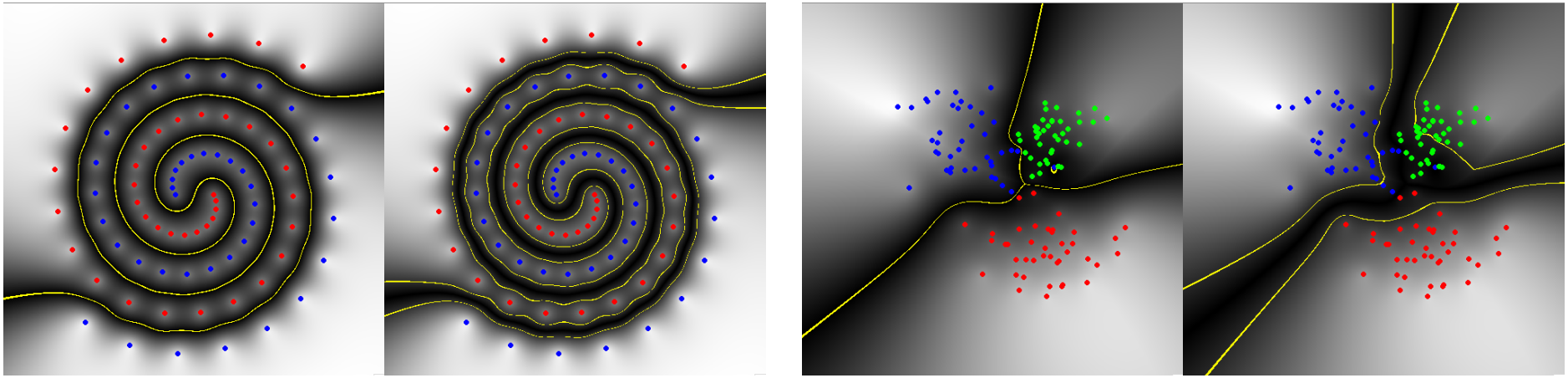
It rather simple to find the borders of the classes:

$$B = \{\mathbf{x} \in \mathcal{X} \text{ s.t. } |\tau_{r(1)}(\mathbf{x}) - \tau_{r(2)}(\mathbf{x})| = 0\}$$

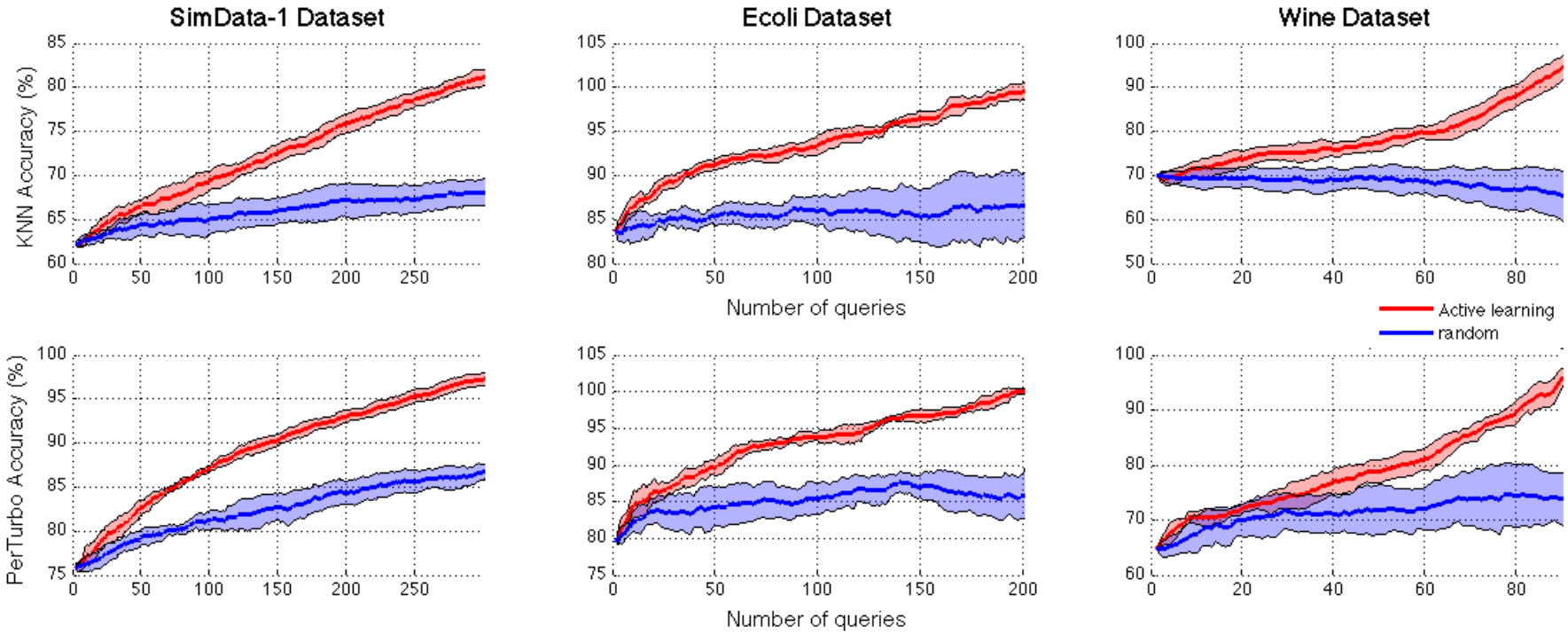
where $r(i)$ is the ith least perturbated classes by x. Then,

$$B' = \{\mathbf{x} \in \mathcal{X} \text{ s.t. } |\tau_{r(1)}(\mathbf{x}) - \tau_{r(2)}(\mathbf{x})| < \gamma\}$$

with $\gamma < 1$, corresponds to the region to query with an active learning policy.



Active Learning (2)



Others

Regularization:

1. What if $\mathbf{K}_\ell$ is not invertible ? $\Rightarrow$ pseudo inverse
2. Matrix regularization

Graph Laplacian Eigenmaps, dimensionality reduction:

1. $\mathbf{K}_\ell$ has the same spectrum as Graph Laplacian
2. We provide a new "projection method" for GLE [Belkin, 2003]

Kernel Mahalanobis distance:

1. Mahalanobis in Kernel space [Haasdonk & Pekalska, 2008]
2. The perturbation measure is similar to some of their distances

Exhaustive study for parameter tuning:

1. The standard deviation σ associated to the Gaussian kernel
2. Additional parameters: Regularization, spectrum truncation...

Outline

PerTurbo: Main results

1. Classification algorithm inspired by computer graphics results
2. With very few parameters to tune
3. Performances similar to SVM
4. Perspectives:
    1. Active learning
    2. Matrix regularization
    3. Manifold learning and dimensionality reduction
    4. Kernel Mahalanobis distance

Questions

Thank you !