Institut Charles Delaunay
université de technologie Troyes

Advanced method for decision
Pattern recognition

One class SVM

- Chapitre 4 -

Outline

Outline

Motivations

Motivations

- Novelty detection
- Quantile estimation

Situations

- Only one kind of observations are available
- Classes without samples
- Sensor failures
- ...

Problem

- A representation space $\mathcal{X}_l \subseteq \mathbb{R}^l$
- A data set of iid observations $A_n = \{x_{1:n}\}$
- A random variable X follows an unknown probability distribution P.

Aim :

Estimate a "simple" subset $S \subset \mathcal{X}_l$ such that $P(X \ni S) = 1 - \mu$ with μ a predefined value.

Main elements of solutions

- Approximate the solution using a function f
- $f(x) > 0$ if $x \in S$
- f is a linear combination of kernels expressed in a transformed space
- f is determined by solving a quadratic problem

Outline

Support and Quantiles

- A data set of iid observations $A_n = \{x_{1:n}\}$ with $x_i \in \mathcal{X}_m$
- The RV X follows an unknown probability distribution P
- $\mathcal{C}$ is the class of the measurable subsets $\mathcal{X}_m$
- λ a measurement function defined on $\mathcal{C}$

Definition :

- $U(\mu)$ is a quantile if :

$$U(\mu) = \inf\{\lambda(C)|P(C) \geq \mu, C \in \mathcal{C}\}$$

Intepretation :
The smallest subset containing a probability mass μ.
Remark :
$U(1)$ is the support of P

Estimation of a quantile

- Empirical estimator :

$$P^n_{emp}(C) = \frac{1}{n}\sum_{i=1}^{n} I_C(x_i)$$

- $C_\lambda(\mu)$, $C^n_\lambda(\mu)$: the C which reach the infimum (greatest lower bound).

Standard measure : the "volume" of C (Lebesgue measure)
We seek C with "minimal" volume
Determine an estimate of minimum volume $C^n_\lambda(\mu)$ is insuffisant :
Ensure ability to generalize.

Quantile estimation

Consequences

- tradeoff between quality of learning - complexity of the learner (VC dimension).
- Restrict or control the set of subsets C eligible.
- With kernel methods : implicit definition of $\mathcal{C}$ via K
- Minimize a quadratic form in the Hilbert space which allows to control the complexity of the function defining C
- We use $\lambda(C_{w,\rho}) = \|w\|^2$ (small VC dimension) with $C_{w,\rho} = \{x | f_w(x) \geq \rho\}$

ν-1SVM

aim :

- Find the best hyperplane that separates the data from the origin
- Maximize the margin
- The value of $f(x)$ depends on the position x with respect to the hyperplane

To find f we solve :

$$\begin{array}{rl} \underset{w \in \mathcal{H}, \xi, \rho}{\text{minimize}} & \frac{1}{2}\|w\|^2 - \rho + \frac{1}{n\nu}\sum_{i=1}^{n} \xi_i \\ \text{with} & \langle w, \Phi(x_i)\rangle \geq \rho - \xi_i \\ \text{and} & \xi_i \geq 0 \quad \forall i = 1 : n \end{array}$$

with $\nu \in]0, 1]$.
$f(x) = signe\left(\langle w, \Phi(x)\rangle - \rho\right)$

ν-1SVM

Resolution :

$$\begin{aligned}\mathfrak{L}(w,\xi,\rho,\alpha,\beta) &= \frac{1}{2}\|w\|^2 + \frac{1}{n\nu}\sum_{i=1}^{n}\xi_i \\ &\quad - \sum_i \alpha_i(\langle w,\Phi(x_i)\rangle - \rho + \xi_i) - \sum_i \beta_i\xi_i - \rho \\ \text{avec} &\quad \alpha_i \geq 0 \quad \beta_i \geq 0\end{aligned}$$

Cancelling derivatives with respect to variables of the primal problem :

- $w = \sum_i \alpha_i \Phi(x_i)$
- $0 \leq \alpha_i \leq \frac{1}{\nu n}$
- $\sum_i \alpha_i = 1$

ν-1SVM

Dual Problem :

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j K(x_i, x_j)$$
$$\text{with} \qquad 0 \leq \alpha_i \leq \frac{1}{\nu n} \text{and} \qquad \sum_i \alpha_i = 1$$

and

$$\rho = \sum_j (\alpha_j K(x_j, x_i)) \text{ quand } \alpha_i \in]0, \frac{1}{\nu n}[$$

due to KKT conditions.

ν-1SVM

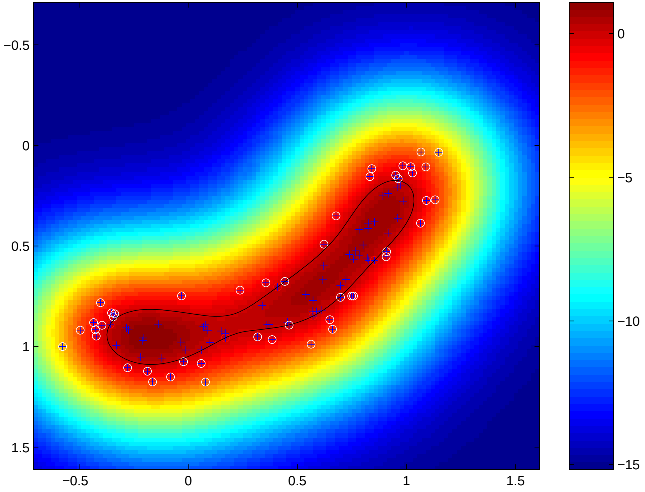


Figure: nu_1svm

ν-SVM

Link with Parzen windows :

- K normalized as a density (Gaussian case)
- $\nu = 1$

Consequence :

- $\alpha_{1:n} = \frac{1}{n}$
- This is the Parzen estimator of a probability density function

Remark :
If K is normalized, $\sum_i \alpha_i K(x_i, x)$ is a probability density that is based solely on the SV.

ν-1SVM

Property 1 :

If the data $\{\Phi(x_{1:n})\}$ are separable then there is only one hyperplane :

- separating all the data from the origin
- maximizing the distance from the origin

This hyperplane is the solution of problem $\rho > 0$:

$$\begin{array}{rl} \underset{w}{\text{minimize}} & \frac{1}{2}\|w\|^2 \\ \text{with} & \langle w, \phi(x_i) \rangle \geq \rho \end{array}$$

ν-SVM

Property 2 :

- Consider the separating hyperplane defined by (w, ρ) for the data $\{(\Phi(x_{1:n}), 1)\}$ then this hyperplane $(w, 0)$ is the optimal hyperplane of the data set $\{(\Phi(x_{1:n}), 1), (-\Phi(x_{1:n}), -1)\}$.

- Consider $(w, 0)$ the optimal hyperplane separating a set of labeled data : $\{(\Phi(x_{1:n}), y_{1:n})\}$ defined such that $\langle w, \phi(x_i)\rangle > 0$ if $y_i = 1$. Moreover, if $\frac{\rho}{\|w\|}$ is the margin of this hyperplane, then (w, ρ) define the support hyperplane for the data set $\{(\Phi(x_{1:n}), 1)\}$.

ν-SVM

Property 3 : If the solution of the 1C-SVM problem is such that $\rho \neq 0$, then

- ν is an upper bound of the proportion of outliers
- ν is a lower bound of the proportion of SV

$$\mathit{PropExclus} \leq \nu \leq \mathit{PropSVs}$$

- If the data are iid according to a continuous distribution function P and the kernel is non-constant then ν is asymptotically equal to the fraction of SVs and excluded.

ν-SVM

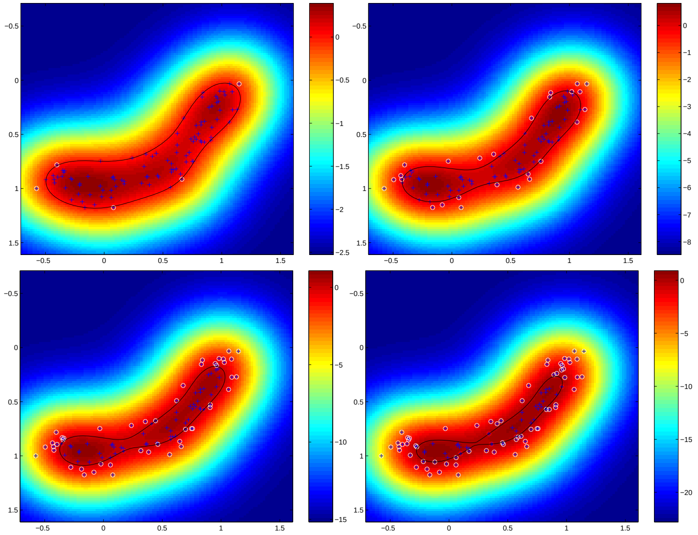


Figure: Examples of ν-1svm with $\nu = 0.1, 0.3, 0.5, 0.7$

SVDD

Aim :

- Find the hypersphere with minimum volume that contains all observations excepted a fraction ν
- The hypersphere is defined by the position of its center a and its radius R

Problem formulation :

$$\begin{array}{cc} \underset{R,\xi}{minimize} & R^2 + \frac{1}{n\nu}\sum_i \xi_i \\ \text{with} & \|a - \Phi(x_i)\|^2 \leq R^2 + \xi_i \\ \text{and} & \xi_i \geq 0 \quad \forall i = 1:n \end{array}$$

Resolution :

$$\begin{array}{rcl} \mathfrak{L}(R,a,\xi) & = & R^2 + \dfrac{1}{n\nu}\displaystyle\sum_i \xi_i - \sum_i(\alpha_i(R^2 + \xi_i - \|a - \Phi(x_i)\|^2)) - \sum_i \beta_i \xi_i \\ \text{with} & & \xi_i \geq 0 \quad \alpha_i \geq 0 \quad \beta_i \geq 0 \quad \forall i = 1:n \end{array}$$

SVDD

Dual Problem :

$$\underset{\alpha}{minimize} \quad \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) - \sum_i \alpha_i K(x_i, x_i)$$

$$\text{with} \qquad 0 \le \alpha_i \le \frac{1}{\nu n} \quad \forall i = 1 : nand \qquad \sum_i \alpha_i = 1$$

Equivalence :

- If $K(x_i, x_i) = Cte$ then SVDD $\Leftrightarrow$ ν-1SVM
- If $K(x, x')$ depends only on $x - x'$ SVDD $\Leftrightarrow$ ν-1SVM

Faire figure

Discriminant 1CSVM

Contexte : It is assumed that we have observations that do not belong to the target class

Problem formulation :

Let n_1 be the number of samples that belong to class ω_1 and n_2 the number of other observations,

$$\begin{array}{rl}
\min\limits_{w \in \mathcal{H}, \xi, \zeta, \rho} & \frac{1}{2}\|w\|^2 - \rho + \frac{1}{n_1 \nu_1} \sum_{i=1}^{n_1} \xi_i + \frac{1}{n_2 \nu_2} \sum_{k=1}^{n_2} \zeta_k \\
\text{with} & \langle w, \Phi(x_i) \rangle \geq \rho - \xi_i \quad \forall i = 1 : n_1 \\
\text{with} & \langle w, \Phi(x_k) \rangle < \rho - \zeta_k \quad \forall k = 1 : n_2 \\
\text{and} & \xi_i \geq 0 \quad \forall i = 1 : n_1 \\
\text{and} & \zeta_k \geq 0 \quad \forall k = 1 : n_2
\end{array}$$

Discriminant 1CSVM

Dual Problem :

$$\begin{array}{rl}
\min\limits_{\alpha} & \sum_{i,j=1}^{n_1} \alpha_i \alpha_j K(x_i, x_j) \\
& -2 \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} \alpha_i \alpha_k K(x_i, x_k) \\
& + \sum_{k,p=1}^{n_2} \alpha_k \alpha_p K(x_k, x_p) \\
\\
\text{with} & \sum_{i=1}^{n_1} \alpha_i - \sum_{k=1}^{n_2} \alpha_k = 1 \\
\text{and} & 0 \leq \alpha_i \leq \frac{1}{n_1 \nu_1} \quad \forall i = 1 : n_1 \\
\text{and} & 0 \leq \alpha_k \leq \frac{1}{n_2 \nu_2} \quad \forall k = 1 : n_2
\end{array}$$

Comments

Outline

SMO Method

Principle :
Divide the primal problem into a sequence of small problems.
Due to equality constraint on $\sum_i \alpha_i \Rightarrow$ it is not possible to change a single value.
At least consider pairs of values.
basic problem : Consider α_i and α_j :

$$\begin{array}{cc} \min\limits_{\alpha_i,\alpha_j} & \frac{1}{2}\left(\alpha_i^2 K_{ii} + \alpha_j^2 K_{jj} + 2\alpha_i\alpha_j K_{ij}\right) + c_i\alpha_i + c_j\alpha_j \\ \text{with} & \alpha_i + \alpha_j = \gamma \\ \text{and} & 0 \leq \alpha_i, \alpha_j \leq \frac{1}{n\nu} \end{array}$$

and

$$\gamma = 1 - \sum_{k\neq i,j} \alpha_k = \alpha_i^{old} + \alpha_j^{old}$$

$$c_i = \sum_{k\neq i,j} \alpha_k K_{ik} \qquad c_j = \sum_{k\neq i,j} \alpha_k K_{jk}$$

decompose in 4 sub-matrix

SMO Method

Resolution : Substituting α_j :

$$\begin{array}{ll} \min\limits_{\alpha_i} & \frac{1}{2}\alpha_i^2\chi + \alpha_i(c_i - c_j + \gamma(K_{ij} - K_{jj})) \\ \text{with} & \chi = K_{ii} - 2K_{ij} + K_{jj} \\ \text{and} & L \leq \alpha_i, \alpha_j \leq H \end{array}$$

and

$$\begin{aligned} L &= \max(0, \gamma - \frac{1}{\nu n}) \\ H &= \min(\frac{1}{\nu n}, \gamma) \end{aligned}$$

where

$$\alpha_i = \min(\max(L, \tilde{\alpha}_i), H)$$

SMO Method

Resolution :

$$
\begin{aligned}
\tilde{\alpha}_i &= \alpha_i^{old} + \frac{c_j - c_i + \alpha_j^{old} K_{jj} - \alpha_i^{old} K_{ii} + K_{ij}(\alpha_i^{old} - \alpha_j^{old})}{\chi} \\
&= \alpha_i^{old} + \frac{f^{old}(x_j) - f^{old}(x_i)}{\chi}
\end{aligned}
$$

Initialization :

Draw $n\nu$ observations and fix values $\alpha_i = \frac{1}{n\nu}$.
If $n\nu$ is not an integer, give the supplement to the last.

SMO Method

SVs Choice :

- We seek a sample that violates the KKT conditions

$$f(x_i)\alpha_i > 0$$

  case $f(x_i) > 0$ and $\alpha_i \neq 0$ or

$$-f(x_i)(\frac{1}{n\nu} - \alpha_i) > 0$$

  case $f(x_i) < 0$ and $\alpha_i \neq \frac{1}{n\nu}$

- Find the sample x_j which maximizes on k $|f(x_i) - f(x_k)|$ for $x_k \in SV$ with $SV = \{x_k | \alpha_k \in]0, \frac{1}{n\nu}[\}$

In practice this operation is repeated on all the data and then several times only on SVs.
If no violation of KKT conditions remains on $\mathcal{A}_n$ then it is finished !

SMO Method

Simple algorithm :

1. Initialize α and $\rho = \max_i f(x_i)$
2. While the KKT conditions are not satisfied for all x_i
    1. Choose α_i and α_j
    2. Initialize α_i^{old} and α_j^{old}
    3. Calculate $\tilde{\alpha}_i$
    4. Determine $\alpha_i = \min(\max(L, \tilde{\alpha}_i), H)$
    5. Deduce α_j
    6. Calculate ρ using the fact that $\rho = f(SV|border)$
3. Return α and ρ

Regularization path

Problem formulation :

$$\begin{array}{rc} \min\limits_{w,\rho,\xi} & \frac{n\nu}{2}\|w\|^2 - n\nu\rho + \sum_{i=1}^{n} \xi_i \\ \text{with} & \langle w, \Phi(x_i)\rangle \geq \rho - \xi_i \quad \forall i = 1:n \\ \text{and} & \xi_i \geq 0 \quad \forall i = 1:n \end{array}$$

Dual Problem :

$$\begin{array}{rc} \text{minimize}_{\alpha} & \frac{1}{2n\nu} \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \\ \text{with} & 0 \leq \alpha_i \leq 1 \quad \forall i = 1:n \\ \text{with} & w = \frac{1}{n\nu} \sum_i \alpha_i \Phi(x_i) \\ \text{and} & \sum_i \alpha_i = n\nu \end{array}$$

Regularization path

Observation 1 :
For a given value ν there is a solution α^ν,
The function $f^\nu(x)$ divides A_n in 3 groups :

- $\mathcal{C} = \{i | f^\nu(x_i) > 0 \text{ et } \alpha_i^\nu = 0\}$
- $\mathcal{M} = \{i | f^\nu(x_i) = 0 \text{ et } \alpha_i^\nu \in]0,1[\}$
- $\mathcal{E} = \{i | f^\nu(x_i) < 0 \text{ et } \alpha_i^\nu = 1\}$

Observation 2 :
When ν changes without change of the groups composition, only the α_i associated with the elements of $\mathcal{M}$ change.
Observation 3 :
When ν changes while the composition of the groups stay the same, values of the α_i associated with the elements of $\mathcal{M}$ change linearly with ν.

Regularization path

Proof :

Given ν^m and ν^M such that the groups stay the same for any $\nu \in [\nu^m, \nu^M]$

Consequence :

$$\sum_{k\in\mathcal{M}} \alpha_k^{\nu^{\boldsymbol{M}}} - \sum_{k\in\mathcal{M}} \alpha_k^{\nu} = n(\nu^M - \nu)$$

We define $g^\nu(x) = \langle w^\nu, \Phi(x)\rangle - \frac{\alpha_{\mathbf{0}}^\nu}{n\nu}$ with $\rho^\nu = \frac{\alpha_{\mathbf{0}}^\nu}{n\nu}$

$$g^\nu(x) = \frac{1}{n\nu}\left(\sum_i \alpha_i^\nu K(x, x_i) - \alpha_0^\nu\right)$$

$$\begin{aligned} g^\nu(x) &= g^\nu(x) - \frac{\nu^M}{\nu} g^{\nu^{\boldsymbol{M}}}(x) + \frac{\nu^M}{\nu} g^{\nu^{\boldsymbol{M}}}(x) \\ &= \frac{1}{n\nu}\left(\sum_i (\alpha_i^\nu - \alpha_i^{\nu^{\boldsymbol{M}}})K(x, x_i) - (\alpha_0^\nu - \alpha_0^{\nu^{\boldsymbol{M}}}) + n\nu^M g^{\nu^{\boldsymbol{M}}}(x)\right) \end{aligned}$$

Regularization path

For any observation $x_k \in \mathcal{M}$ we have :
$g^{\nu^{\boldsymbol{M}}}(x_k) = g^{\nu}(x_k) = 0$

$$\sum_{l \in \mathcal{M}} (\alpha_l^{\nu} - \alpha_l^{\nu^{\boldsymbol{M}}}) K(x_k, x_l) - (\alpha_0^{\nu} - \alpha_0^{\nu^{\boldsymbol{M}}}) = 0$$

$$\begin{cases} K\left(\alpha^{\nu} - \alpha^{\nu^{\boldsymbol{M}}}\right) - \left(\alpha_0^{\nu} - \alpha_0^{\nu^{\boldsymbol{M}}}\right)\mathbf{1} & = & \mathbf{0} \\ \mathbf{1}^T\left(\alpha^{\nu} - \alpha^{\nu^{\boldsymbol{M}}}\right) & = & n\nu - n\nu^{M} \end{cases}$$

Let :

$$\begin{aligned} A &= \begin{pmatrix} K & -\mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \\ c^T &= [0 \ldots 0, 1] \end{aligned}$$

Regularization path

Thus :

$$A\left(\begin{pmatrix}\alpha^{\nu}\\ \alpha_0^{\nu}\end{pmatrix}-\begin{pmatrix}\alpha^{\nu^{\boldsymbol{M}}}\\ \alpha_0^{\nu^{\boldsymbol{M}}}\end{pmatrix}\right)=(n\nu-n\nu^{M})c \qquad (1)$$

Thus :

$$\begin{pmatrix}\alpha^{\nu}\\ \alpha_0^{\nu}\end{pmatrix}=\begin{pmatrix}\alpha^{\nu^{\boldsymbol{M}}}\\ \alpha_0^{\nu^{\boldsymbol{M}}}\end{pmatrix}+(n\nu-n\nu^{M})A^{-1}c \qquad (2)$$

Regularization path

Group change :

- x_k de $\mathcal{M}$ vers $\mathcal{E}$ $(\alpha_k \to 1)$

$$\nu = \frac{1 - \alpha_k^{\nu^{\boldsymbol{M}}}}{n(A^{-1}c)_k} + \nu^M$$

- x_k de $\mathcal{M}$ vers $\mathcal{C}$ $(\alpha_k \to 0)$

$$\nu = \frac{-\alpha_k^{\nu^{\boldsymbol{M}}}}{n(A^{-1}c)_k} + \nu^M$$

- x_k vers $\mathcal{M}$ (eq 2)

$$\nu = \nu^M \left(\frac{[K(x_k, .), -1](A^{-1}c) - g^{\nu^{\boldsymbol{M}}}(x_k)}{[K(x_k, .), -1](A^{-1}c)} \right)$$

Regularization path

Group change case 3 :

$$g^{\nu}(x_k) = \frac{1}{n\nu}\left(\sum_i (\alpha_i^{\nu} - \alpha_i^{\nu^{\boldsymbol{M}}})K(x_k, x_i) - (\alpha_0^{\nu} - \alpha_0^{\nu^{\boldsymbol{M}}}) + n\nu^M g^{\nu^{\boldsymbol{M}}}(x_k)\right) = 0$$

Thus :

$$\sum_i (\alpha_i^{\nu} - \alpha_i^{\nu^{\boldsymbol{M}}})K(x_k, x_i) - (\alpha_0^{\nu} - \alpha_0^{\nu^{\boldsymbol{M}}}) = -n\nu^M g^{\nu^{\boldsymbol{M}}}(x_k)$$

Thus :

$$[K(x_k, .), -1]\left(\begin{pmatrix} \alpha^{\nu} \\ \alpha_0^{\nu} \end{pmatrix} - \begin{pmatrix} \alpha^{\nu^{\boldsymbol{M}}} \\ \alpha_0^{\nu^{\boldsymbol{M}}} \end{pmatrix}\right) = -n\nu^M g^{\nu^{\boldsymbol{M}}}(x_k)$$

Using eq. 2 :

$$n[K(x_k, .), -1]A^{-1}c(\nu - \nu^M) = -n\nu^M g^{\nu^{\boldsymbol{M}}}(x_k)$$

Regularization path

Step :
$\nu^m = \max(\nu_{changement})$
Replace ν^M by the value ν^m
progression by decreasing value of ν

Stop :
When $\mathcal{E}$ is empty !

Regularization path

Initialization :

Start with $\nu = 1 - \frac{\epsilon}{n}$

A single x_k is in the margin, the others are outliers ($\alpha_i = 1$ et ($\alpha_k = 1 - \epsilon$). $\rho = f(x_k)$

To find k :

$$\min_k \frac{1}{2n\nu} \sum_{i,j} \alpha_i \alpha_j K_{ij} - \frac{\epsilon}{n\nu} \sum_i \alpha_i K_{ik} + \frac{\epsilon^2}{2n\nu} K_{kk}$$

Thus :

$$\min_k \frac{1}{2n\nu} \sum_{i,j} K_{ij} - \frac{\epsilon}{n\nu} \sum_i K_{ik} + \frac{\epsilon^2}{2n\nu} K_{kk}$$

Which means choosing the closest x_k to the barycenter in the Hilbert space :

$$\max_k \sum_{i \neq k} K_{ik}$$

If several x_k enter simultaneously the margin ($P = 0$) : solve the optimisation problem with ν close to 1.

Outline

PCA

Principle

Projection from $\mathbb{R}^l$ to $\mathbb{R}^d$
Minimizing the deformation of the sample cloud
Retain most of the variance
Decision based on reconstruction error.

PCA Method

We search u such that $\|u\| = 1$ and the projection of X on the axis carried by u captures the most variance.

$$\begin{array}{ll} & \max_{u} \quad u^t X^t X u \\ \text{avec} & \quad u^t u = 1 \end{array}$$

Dual Problem :

$$\max_{\lambda} \quad \min_{u} \quad - u^t X^t X u + \lambda (u^t u - 1)$$

PCA

Method - continuation

We search v such that $\|v\| = 1$ and $v \perp u$ and the projection of X on the axis carried by v captures the most variance in the subspace.

$$\max_{v} \quad v^t X^t X v$$
$$\text{avec} \quad v^t v = 1 \text{ et } v^t u = 0$$

Dual problem :

$$\max_{\lambda,\beta} \quad \min_{v} \quad -v^t X^t X v + \lambda(v^t v - 1) + \beta v^t u$$

Solution

Eigenvectors
Eigenvalues

KPCA

Principle

Perform a PCA in a transformed space $\mathcal{H}$

Given Φ a transformation such that :

$$\Phi \quad : \quad \mathbb{R}^l \rightarrow \mathcal{H}$$
$$x \mapsto \Phi(x)$$

Hypothesis : $\sum_i \tilde{\Phi}(x_i) = 0$
One defines $\tilde{S}_f = \frac{1}{n} \sum_i \tilde{\Phi}(x_i)\tilde{\Phi}(x_i)^t$

We search λ and V such that :

$$\tilde{S}_f V = \lambda V$$

KPCA

Solution

V a vector of the space $\{\tilde{\Phi}(x_1), \ldots, \tilde{\Phi}(x_n)\} \Rightarrow$

$$\exists \alpha \text{ st. } V = \sum_i \alpha_i \tilde{\Phi}(x_i)$$

with $\tilde{K}_{i,j} =< \tilde{\Phi}(x_i), \tilde{\Phi}(x_j) >$ Consequences :

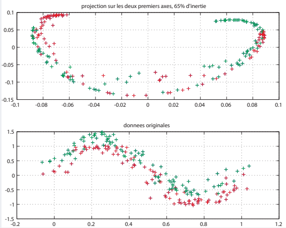
$$K\alpha = n\lambda\alpha$$

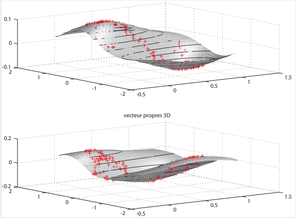
α are the eigenvectors of K
$KPCA_d(x) = \sum_i \alpha_i^d \tilde{K}(x_i, x)$

KPCA

Interpretation

- Maximize the variance in $\mathcal{H}$
- Minimize the reconstruction error in $\mathcal{H}$
- Minimize the representation entropy
- Maximize mutual information in relation with the data
- If Φ is a polynomial, relation with high order moments

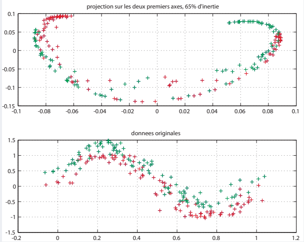


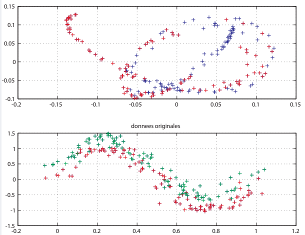


KPCA

Interpretation

- Maximize the variance in $\mathcal{H}$
- Minimize the reconstruction error in $\mathcal{H}$
- Minimize the representation entropy
- Maximize mutual information in relation with the data
- If Φ is a polynomial, relation with high order moments

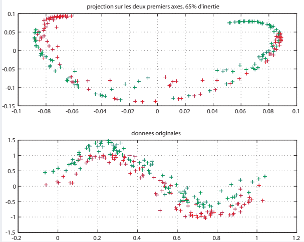


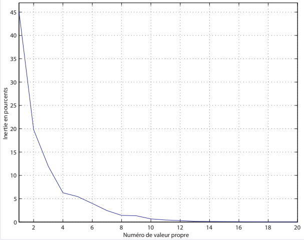


KPCA

Interpretation

- Maximize the variance in $\mathcal{H}$
- Minimize the reconstruction error in $\mathcal{H}$
- Minimize the representation entropy
- Maximize mutual information in relation with the data
- If Φ is a polynomial, relation with high order moments





Outline

Conclusion

Contributions

- Detection
- Classification