

Détection d'anomalies avec retour utilisateur

Julien LESOUPLE

TéSA

24 Février 2020

En collaboration avec :

Cédric BAUDOUIN, Marc SPIGAI, Andrés TROYA-GALVIS (TAS)
Jean-Yves TOURNERET (IRIT-ENSEEIH)



Plan

Contexte

Machines à vecteurs supports

Détection d'anomalies

Ajout de retour utilisateur

Conclusion



Plan

Contexte

Machines à vecteurs supports

Détection d'anomalies

Ajout de retour utilisateur

Conclusion



Introduction (1/2)

Intérêt grandissant pour apprentissage automatique et IA

- ▶ évolution des ordinateurs (calculs, mémoire)
- ▶ quantité de données (de plus en plus de capteurs, IOT)

Principales utilisations

- ▶ régression : prediction d'une valeur continue
 - ▶ prédire le prix d'une maison en connaissant sa surface, le nombre de pièces, ...
- ▶ classification : prediction d'une valeur discrète
 - ▶ prédire si un patient est malade en connaissant son ECG, son poids, ...



Introduction (2/2)

Généralement deux étapes

- ▶ étape d'apprentissage
 - ▶ des données d'entraînement sont fournies à l'algorithme pour apprendre la fonction de prédiction
- ▶ étape de validation
 - ▶ des données de test sont fournies à la fonction de prédiction pour valider les résultats

Deux manières d'effectuer l'apprentissage

- ▶ Supervisé
 - ▶ les sorties attendues sont connues et fournies à l'algorithme
- ▶ Non-supervisé
 - ▶ les sorties attendues sont inconnues



Détection d'anomalies¹

Principe de la détection d'anomalies (AD)

- ▶ classification à deux classes : données normales et anomalies
- ▶ en général : apprentissage non-supervisé

Particularités

- ▶ les jeux de données sont déséquilibrés (très peu d'anomalies)
- ▶ parfois, pas d'anomalies dans les données d'apprentissage

Approche classique pendant l'apprentissage

- ▶ on suppose que la majorité des données appartient à la classe nominale
- ▶ les données nominales sont contenus dans une petit région de l'espace
- ▶ les anomalies sont des données loin de cette région

1. Varun CHANDOLA, Arindam BANERJEE et Vipin KUMAR. "Anomaly Detection : A Survey". In : *ACM Computing Surveys* 41.3 (2009), 15 :1–15 :58.

Contexte

Machines à vecteurs supports

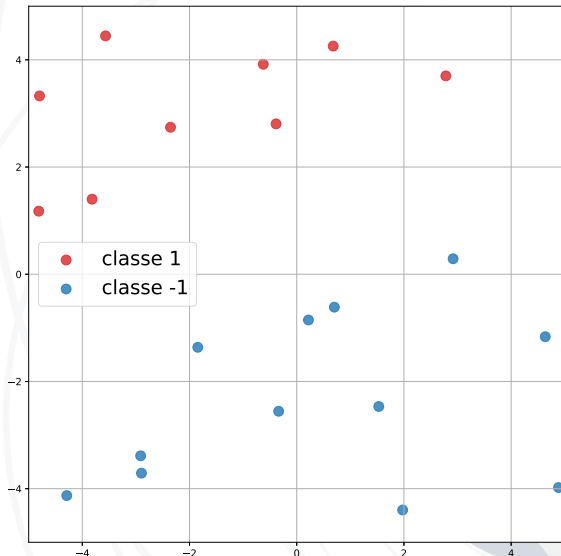
Détection d'anomalies

Ajout de retour utilisateur

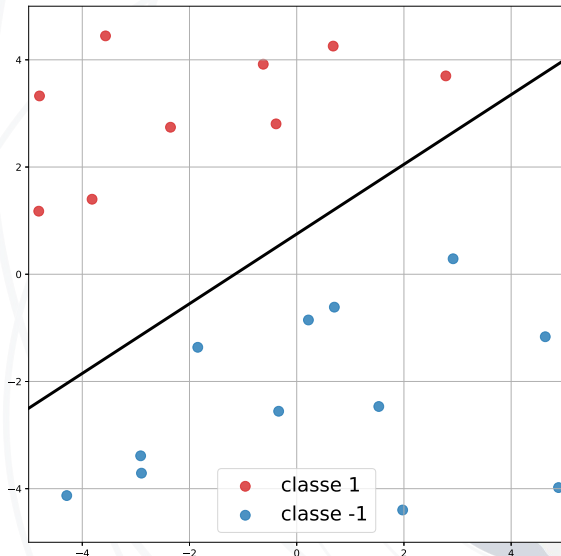
Conclusion



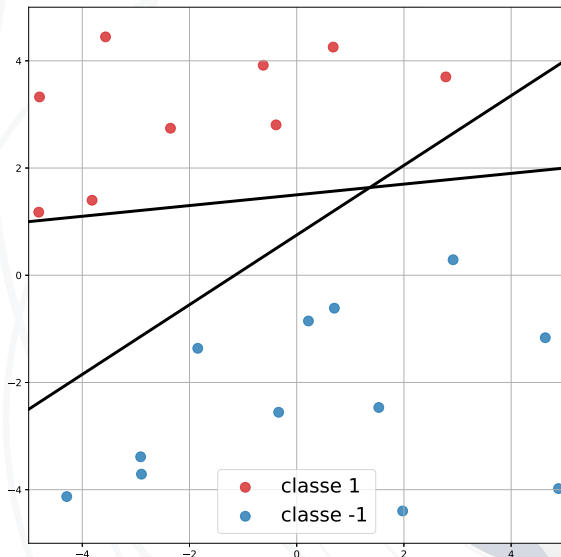
Exemple classification supervisée



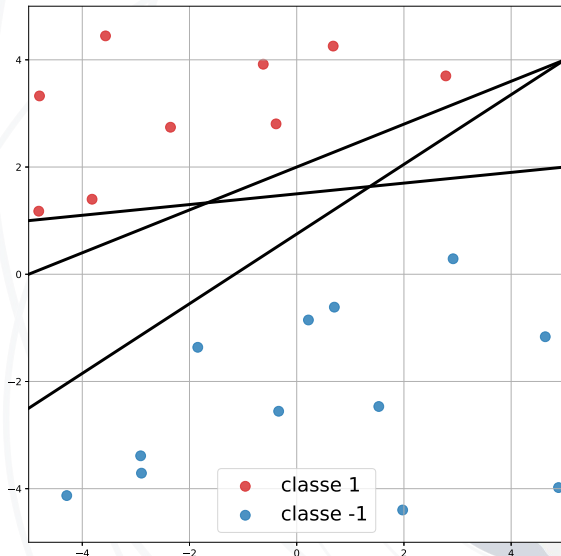
Exemple classification supervisée



Exemple classification supervisée



Exemple classification supervisée

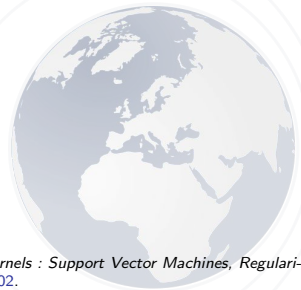


Machine à vecteurs supports pour classification binaire^{2 3}

Données d'entraînements labellisées $(\mathbf{x}_i, y_i)$, $y_i = \pm 1$

Equation d'un hyperplan : $\mathbf{w}^T \mathbf{x} + b = 0$

Règle de classification : $y = \text{sign}(\mathbf{w}^T \mathbf{x} + b)$



2. [Bernhard SCHÖLKOPF](#) et [Alexander J. SMOLA](#). *Learning with Kernels : Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA : The MIT Press, 2002.

3. [Christopher M. BISHOP](#). *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.

Machine à vecteurs supports pour classification binaire^{2 3}

Données d'entraînements labellisées $(\mathbf{x}_i, y_i)$, $y_i = \pm 1$

Equation d'un hyperplan : $\mathbf{w}^T \mathbf{x} + b = 0$

Règle de classification : $y = \text{sign}(\mathbf{w}^T \mathbf{x} + b)$

Principe : hyperplan qui maximise la distance minimale d'une classe à l'hyperplan

$$\begin{aligned} \arg \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|_2^2 \\ \text{s.t. } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \end{aligned}$$

2. [Bernhard SCHÖLKOPF](#) et [Alexander J. SMOLA](#). *Learning with Kernels : Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA : The MIT Press, 2002.

3. [Christopher M. BISHOP](#). *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.

Résolution (1/2)

Lagrangien

$$\mathcal{L}(\mathbf{w}, b) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{i=1}^n \alpha_i [1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)]$$

Conditions de Karush Kuhn Tucker

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i = \mathbf{X}^T \mathbf{Y} \boldsymbol{\alpha}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^n y_i \alpha_i = 0$$

$$\alpha_i [1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)] = 0$$

$$\alpha_i \geq 0$$

avec $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$, $\mathbf{Y} = \text{diag}(y_i)$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$



Résolution (2/2)

Problème dual

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \alpha^T \mathbf{Y} \mathbf{X} \mathbf{X}^T \mathbf{Y} \alpha - \mathbf{1}^T \alpha \\ \text{s.t. } & 0 \leq \alpha_i \\ & \sum_{i=1}^n y_i \alpha_i = 0 \end{aligned}$$

Solution

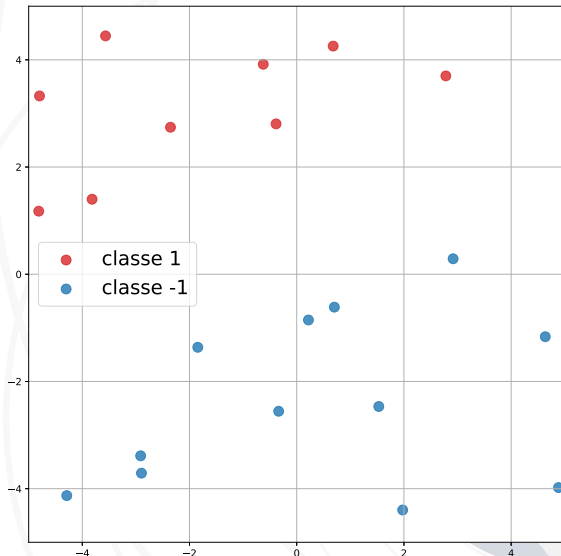
$$\alpha_i [1 - y_i (\mathbf{w}^T \mathbf{x}_i + b)] = 0 \Rightarrow \begin{cases} 1 = y_i (\mathbf{w}^T \mathbf{x}_i + b) & \text{si } \alpha_i > 0 \\ y_i (\mathbf{w}^T \mathbf{x}_i + b) > 1 & \text{si } \alpha_i = 0 \end{cases}$$

$$\mathbf{w} = \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i = \sum_{\alpha_i \neq 0} y_i \alpha_i \mathbf{x}_i$$

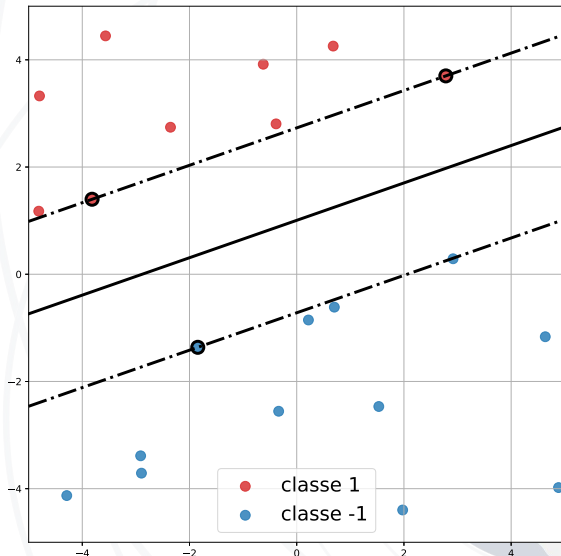
Prédiction de la classe associée à $\mathbf{x}$

$$\hat{y} = \text{sign}(\mathbf{w}^T \mathbf{x} + b) = \text{sign}(\alpha^T \mathbf{Y} \mathbf{X} \mathbf{x} + b)$$

Exemple classification supervisée



Exemple classification supervisée



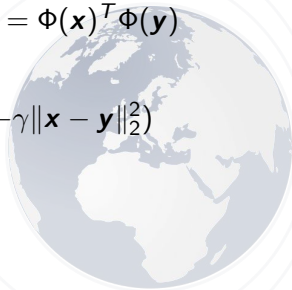
Données non séparables linéairement

Idée : trouver une fonction Φ telle que les $\Phi(\mathbf{x}_i)$ sont linéairement séparables

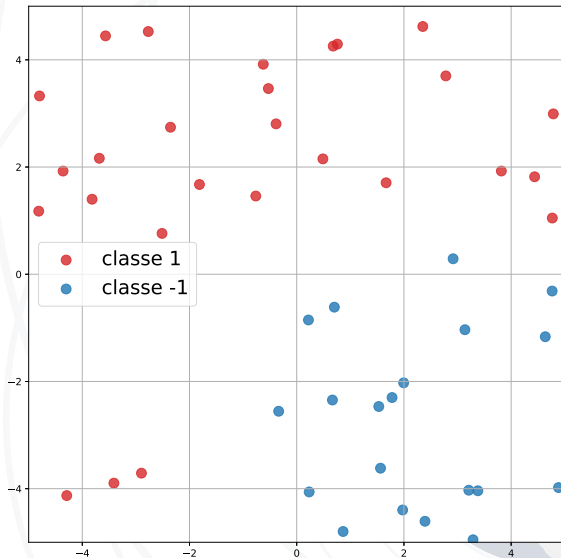
Astuce : la fonction de décision dépend seulement des $\Phi(\mathbf{x})^T \Phi(\mathbf{x}_i)$

→ utilisation d'un noyau $k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$

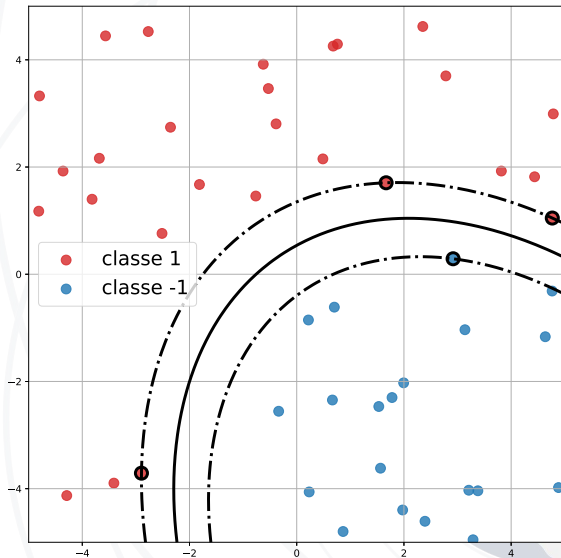
Exemple : noyau Gaussien $k(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|_2^2)$



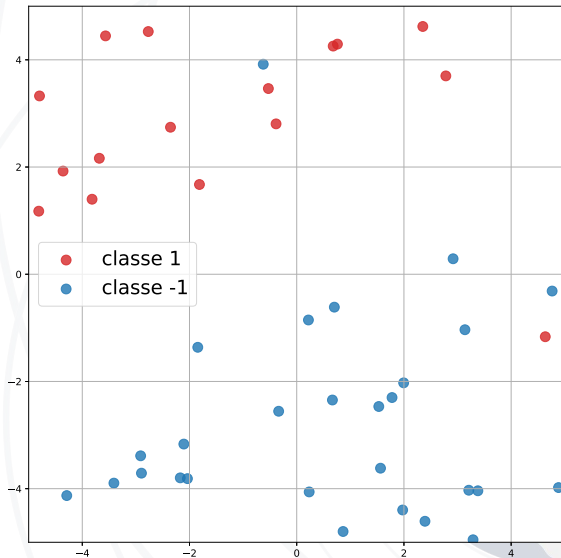
Illustration



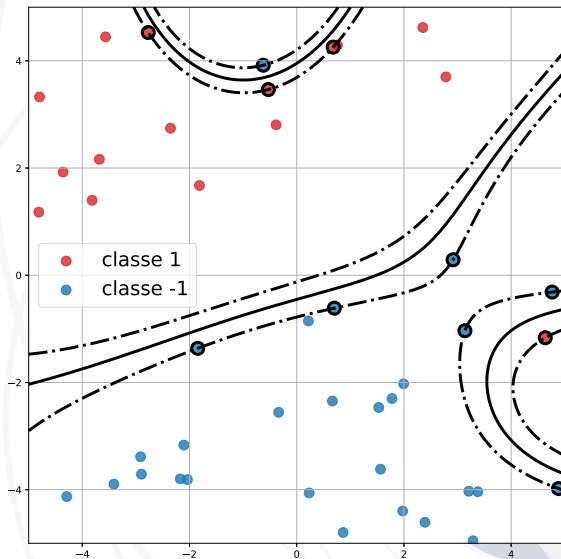
Illustration



Autre cas



Autre cas



But : autoriser certaines données à être du mauvais côté de la frontière de décision

Principe : introduire des variables de relâchement ξ_i non nulles pour une partie des données

$$\arg \min_{\mathbf{w}, b, \xi, \rho} \frac{1}{2} \|\mathbf{w}\|_2^2 - \gamma \nu + \frac{1}{n} \sum_{i=1}^n \xi_i$$

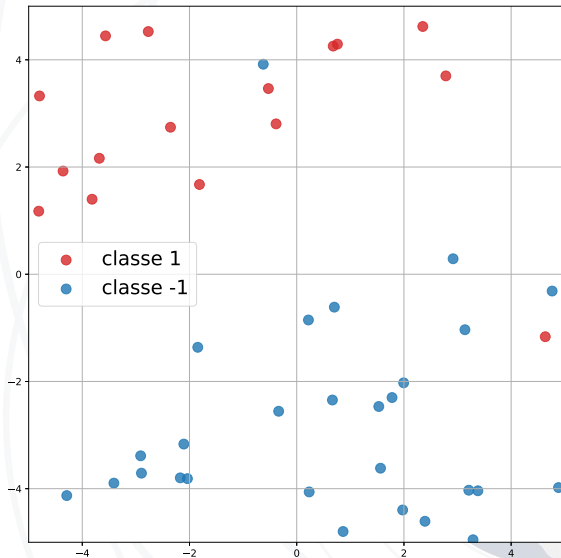
$$\text{s.t. } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq \gamma - \xi_i$$

$$\xi_i \geq 0$$

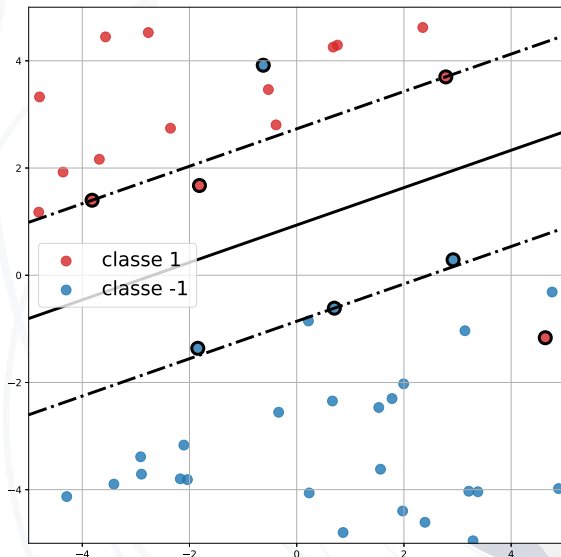
$$\gamma \geq 0$$

4. Bernhard SCHÖLKOPF, Alex J. SMOLA, Robert C. WILLIAMSON et Peter L. BARTLETT. "New Support Vector Algorithms". In : *Neural Computation* 12.5 (2000), p. 1207–1245.

Illustration



Illustration



Plan

Contexte

Machines à vecteurs supports

Détection d'anomalies

Ajout de retour utilisateur

Conclusion



Approche classique pendant l'apprentissage

- ▶ on suppose que la majorité des données appartient à la classe nominale
- ▶ les données nominales sont contenues dans une petite région de l'espace
- ▶ les anomalies sont des données éloignées de cette région



Application à la détection d'anomalies : One-class SVM⁵

Principe

$$\begin{aligned} \arg \min_{\mathbf{w}, b, \xi, \gamma} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{n} \sum_{i=1}^n \xi_i - \gamma \nu \\ \text{s.t. } & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq \gamma - \xi_i \\ & \xi_i \geq 0 \\ & \gamma \geq 0 \end{aligned}$$


5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWE-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Principe

- ▶ classifieur SVM à une classe : la classe nominale ($y_i = 1$)

$$\begin{aligned} \arg \min_{\mathbf{w}, b, \xi, \gamma} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{n} \sum_{i=1}^n \xi_i - \gamma \nu \\ \text{s.t. } & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq \gamma - \xi_i \\ & \xi_i \geq 0 \\ & \gamma \geq 0 \end{aligned}$$

5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWE-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Application à la détection d'anomalies : One-class SVM⁵

Principe

- ▶ classifieur SVM à une classe : la classe nominale ($y_i = 1$)

$$\begin{aligned} \arg \min_{\mathbf{w}, b, \xi, \gamma} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{n} \sum_{i=1}^n \xi_i - \gamma \nu \\ \text{s.t. } & \mathbf{w}^T \mathbf{x}_i + b \geq \gamma - \xi_i \\ & \xi_i \geq 0 \\ & \gamma \geq 0 \end{aligned}$$

5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWE-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Application à la détection d'anomalies : One-class SVM⁵

Principe

- ▶ classifieur SVM à une classe : la classe nominale ($y_i = 1$)
- ▶ pas de marge de classification

$$\begin{aligned} \arg \min_{\mathbf{w}, b, \xi, \gamma} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{n} \sum_{i=1}^n \xi_i - \gamma \nu \\ \text{s.t. } & \mathbf{w}^T \mathbf{x}_i + b \geq \gamma - \xi_i \\ & \xi_i \geq 0 \\ & \gamma \geq 0 \end{aligned}$$

5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWE-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Application à la détection d'anomalies : One-class SVM⁵

Principe

- ▶ classifieur SVM à une classe : la classe nominale ($y_i = 1$)
- ▶ pas de marge de classification

$$\begin{aligned} \arg \min_{\mathbf{w}, b, \xi} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{n} \sum_{i=1}^n \xi_i \\ \text{s.t. } & \mathbf{w}^T \mathbf{x}_i + b \geq -\xi_i \\ & \xi_i \geq 0 \end{aligned}$$



5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWÉ-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Application à la détection d'anomalies : One-class SVM⁵

Principe

- ▶ classifieur SVM à une classe : la classe nominale ($y_i = 1$)
- ▶ pas de marge de classification

$$\begin{aligned} \arg \min_{\mathbf{w}, \rho, \xi} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{n} \sum_{i=1}^n \xi_i \\ \text{s.t. } & \mathbf{w}^T \mathbf{x}_i \geq \rho - \xi_i \\ & \xi_i \geq 0 \end{aligned}$$



5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWÉ-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Principe

- ▶ classifieur SVM à une classe : la classe nominale ($y_i = 1$)
- ▶ pas de marge de classification
- ▶ on autorise une proportion maximale ν des données n'appartenant pas à la classe nominale

$$\begin{aligned} \arg \min_{\mathbf{w}, \rho, \xi} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{n} \sum_{i=1}^n \xi_i \\ \text{s.t. } & \mathbf{w}^T \mathbf{x}_i \geq \rho - \xi_i \\ & \xi_i \geq 0 \end{aligned}$$



5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWE-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Application à la détection d'anomalies : One-class SVM⁵

Principe

- ▶ classifieur SVM à une classe : la classe nominale ($y_i = 1$)
- ▶ pas de marge de classification
- ▶ on autorise une proportion maximale ν des données n'appartenant pas à la classe nominale

$$\begin{aligned} \arg \min_{\mathbf{w}, \rho, \xi} & \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{1}{\nu n} \sum_{i=1}^n \xi_i - \rho \\ \text{s.t. } & \mathbf{w}^T \mathbf{x}_i \geq \rho - \xi_i \\ & \xi_i \geq 0 \end{aligned}$$

5. Bernard SCHÖLKOPF, John C. PLATT, John C. SHAWE-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON.
"Estimating the Support of a High-Dimensional Distribution". In : *Neural Computation* 13.7 (2001), p. 1443–1471.

Détection d'anomalies par One-Class SVM

Règle de décision (+1 : donnée normale, -1 : anomalie)

$$\hat{y}(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x} - \rho)$$

Score d'anomalie (distance relative à l'hyperplan)

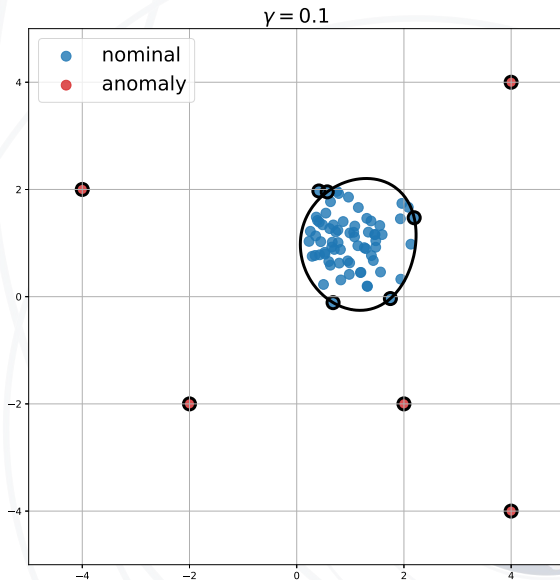
$$s(\mathbf{x}) = -\mathbf{w}^T \mathbf{x}$$



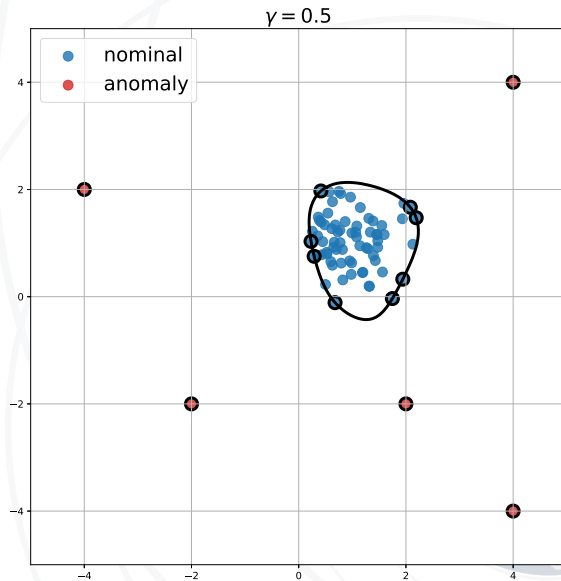
A scatter plot illustrating the detection of anomalies. The plot shows a cluster of nominal data points (blue dots) centered around (1, 1) and several anomaly points (red dots) located at the corners of the plot area. The legend indicates that blue dots represent 'nominal' data and red dots represent 'anomaly' data. The plot is overlaid on a background image of a globe.



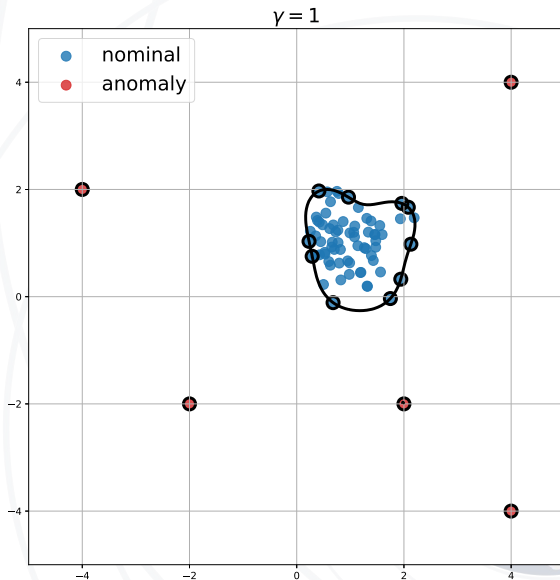
Illustration



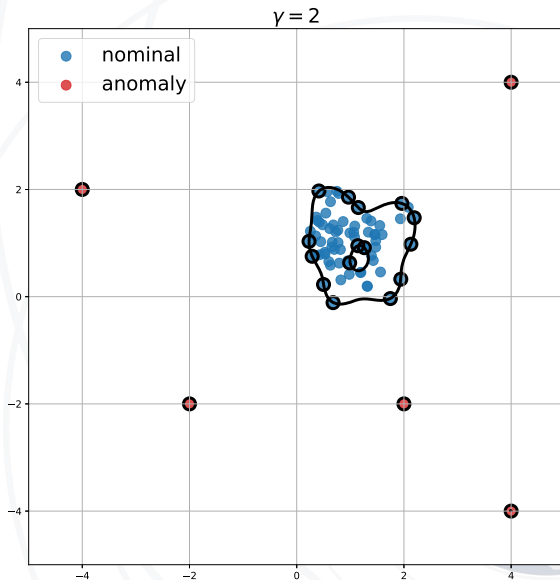
Illustration



Illustration



Illustration



Choix du paramètre du noyau⁶

Noyau Gaussien

$$k(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|_2^2)$$

Parallèle avec la loi Gaussienne

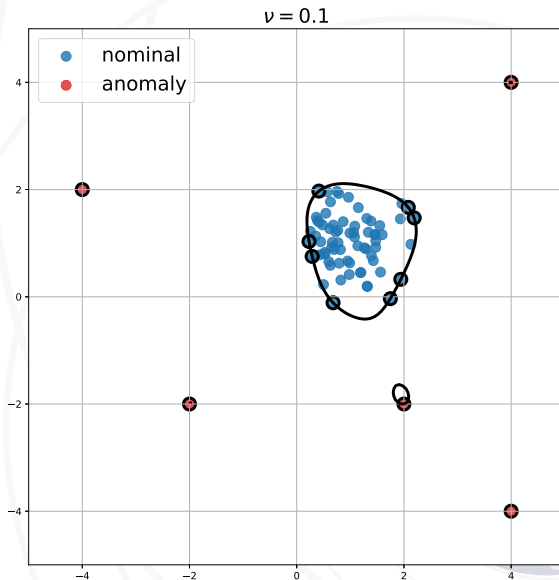
$$\gamma = \frac{1}{2\sigma^2}$$

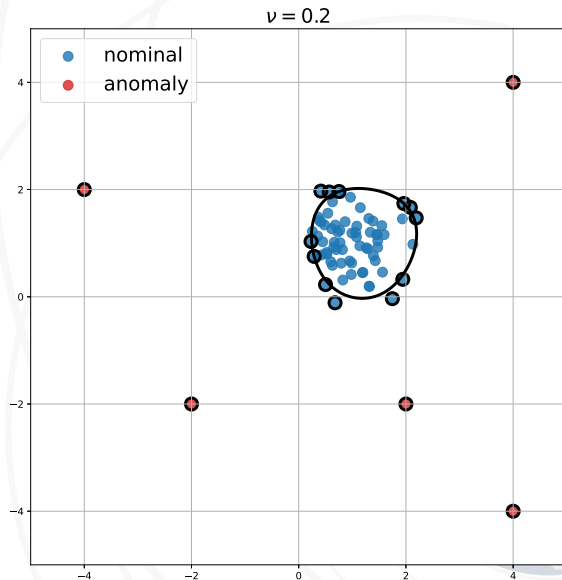
Estimateur robuste de l'écart-type

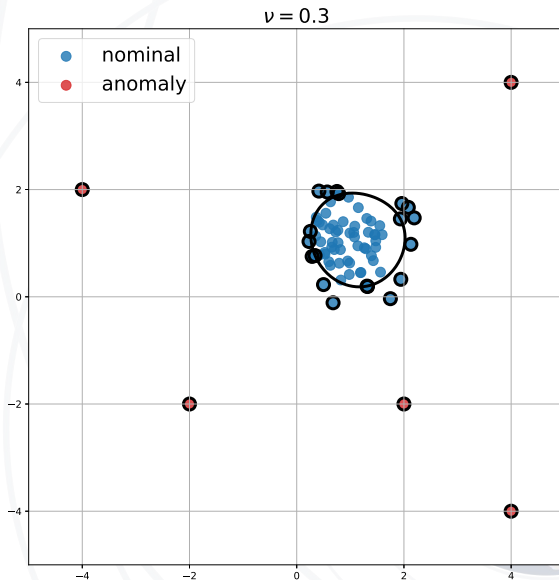
$$\hat{\sigma} = \text{median}(\|\mathbf{x}_i - \mathbf{x}_j\|_2)$$

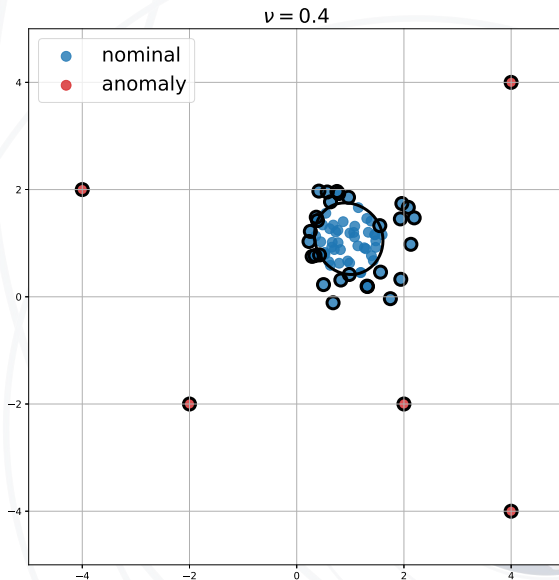


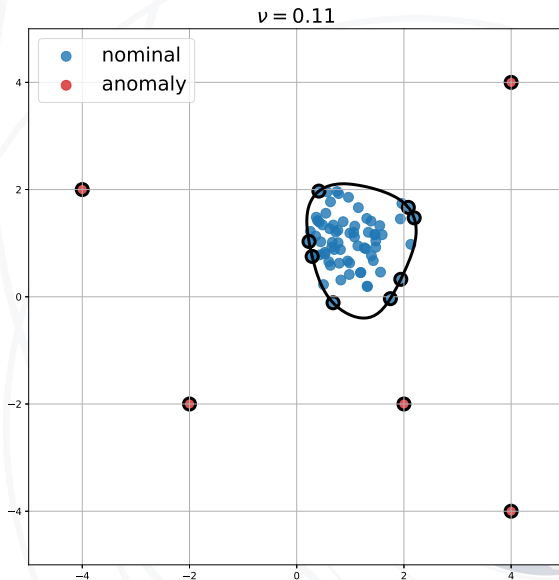
6. Tommi JAAKKOLA, Mark DIEKHANS et David HAUSSLER. "A Discriminative Framework for Detecting Remote Protein Homologies". In : *Journal of Computational Biology* 7.1-2 (2000), p. 95–114.

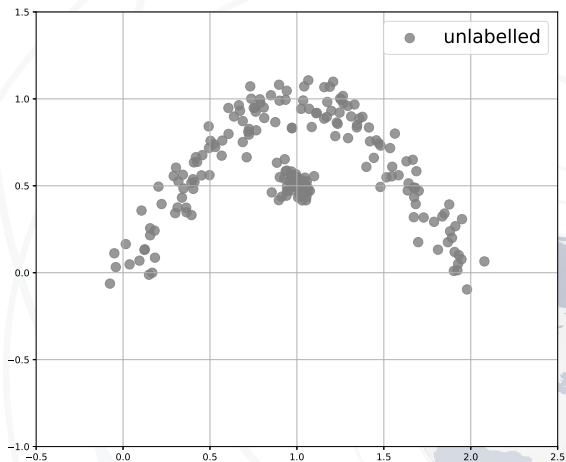


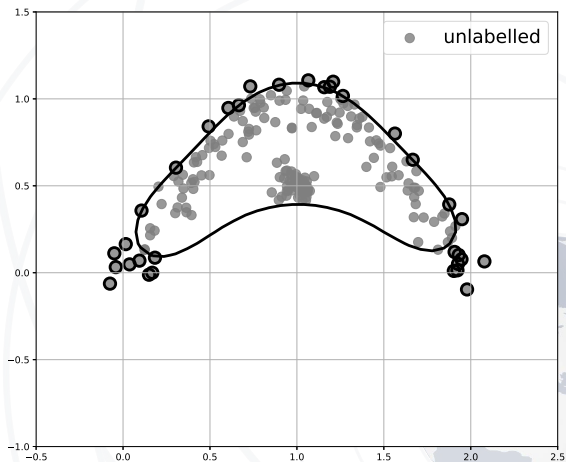


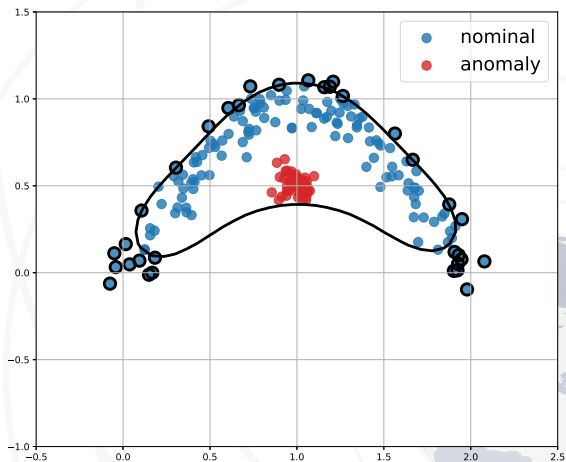


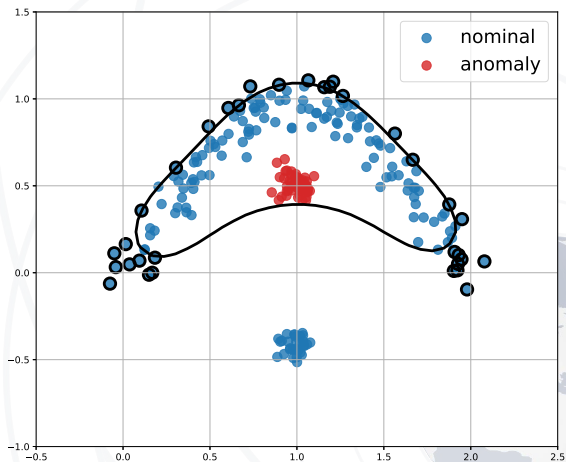












Plan

Contexte

Machines à vecteurs supports

Détection d'anomalies

Ajout de retour utilisateur

Conclusion



Principe

- ▶ étudier tout le jeu d'entraînement est fastidieux
- ▶ les anomalies sont très rares
 - ▶ parfois non-observées pendant l'entraînement

Idée

- ▶ montrer les données les plus ambiguës à un expert pour avoir un label



But

- ▶ montrer le plus d'anomalies possible à l'expert

Algorithme itératif

- ▶ initialisation avec un algorithme de détection d'anomalies non-supervisé (IF⁷ par exemple)
- ▶ boucler jusqu'à un nombre max de requêtes à l'expert
 - ▶ calcul des scores d'anomalies pour les données non labellisées
 - ▶ demande d'un label pour la donnée avec le plus haut score
 - ▶ calcul de la nouvelle frontière de décision

7. Shubhomoy DAS, Weng-Keen WONG, Alan FERN, Thomas G. DIETTERICH et Md Amran SIDDIQUI. "Incorporating Feedback into Tree-based Anomaly Detection". In : *Proc. Workshop on Interactive Data Exploration and Analytics (IDEA)*. Halifax, Canada, 2017, p. 25–33.

8. Shubhomoy DAS, Weng-Keen WONG, Thomas G. DIETTERICH, Alan FERN et Andrew EMMOTT. "Incorporating Expert Feedback into Active Anomaly Discovery". In : *Proc. Int. Conf. on Data Mining (ICDM)*. Barcelona, Spain, 2016, p. 853–858.

But

- ▶ montrer le plus d'anomalies possible à l'expert

Algorithme itératif

- ▶ initialisation avec un algorithme de détection d'anomalies non-supervisé (IF⁷ par exemple)
- ▶ boucler jusqu'à un nombre max de requêtes à l'expert
 - ▶ calcul des scores d'anomalies pour les données non labellisées
 - ▶ demande d'un label pour la donnée avec le plus haut score
 - ▶ calcul de la nouvelle frontière de décision

7. Shubhomoy DAS, Weng-Keen WONG, Alan FERN, Thomas G. DIETTERICH et Md Amran SIDDIQUI. "Incorporating Feedback into Tree-based Anomaly Detection". In : *Proc. Workshop on Interactive Data Exploration and Analytics (IDEA)*. Halifax, Canada, 2017, p. 25–33.

8. Shubhomoy DAS, Weng-Keen WONG, Thomas G. DIETTERICH, Alan FERN et Andrew EMMOTT. "Incorporating Expert Feedback into Active Anomaly Discovery". In : *Proc. Int. Conf. on Data Mining (ICDM)*. Barcelona, Spain, 2016, p. 853–858.

Pour AAD avec IF

- ▶ fonction coût sur-mesure qui prend en compte les retours

Proposition

- ▶ One-class SVM avec une partie des données labellisées

$$\arg \min_{\mathbf{w}, \rho, \xi} \frac{1}{2} \|\mathbf{w}\|_2^2 - \rho + \frac{1}{\nu_1 n_1} \sum_{i=1}^{n_1} \xi_i$$

s.t. $\mathbf{w}^T \Phi(\mathbf{x}_i) - \rho \geq -\xi_i$ and $\xi_i \geq 0$ données non-labellisées

9. Julien LESOUPLE et Jean-Yves TOURNERET. "Incorporating User Feedback Into One-Class Support Vector Machines for Anomaly Detection". In : *Proc. Eur. Conf. on Signal Processing (EUSIPCO)*. Amsterdam, Netherlands, 2020, p. 1608–1612.

Pour AAD avec IF

- ▶ fonction coût sur-mesure qui prend en compte les retours

Proposition

- ▶ One-class SVM avec une partie des données labellisées

$$\arg \min_{\mathbf{w}, \rho, \xi} \frac{1}{2} \|\mathbf{w}\|_2^2 - \rho + \frac{1}{\nu_1 n_1} \sum_{i=1}^{n_1} \xi_i + \frac{1}{\nu_2 n_2} \sum_{l=1}^{n_2} \zeta_l$$

s.t. $\mathbf{w}^T \Phi(\mathbf{x}_i) - \rho \geq -\xi_i$ and $\xi_i \geq 0$ données non-labellisées
 $\mathbf{w}^T \Phi(\mathbf{y}_l) - \rho \geq -\zeta_l$ and $\zeta_l \geq 0$ données normales

9. Julien LESOUPLE et Jean-Yves TOURNERET. "Incorporating User Feedback Into One-Class Support Vector Machines for Anomaly Detection". In : *Proc. Eur. Conf. on Signal Processing (EUSIPCO)*. Amsterdam, Netherlands, 2020, p. 1608–1612.

Pour AAD avec IF

- ▶ fonction coût sur-mesure qui prend en compte les retours

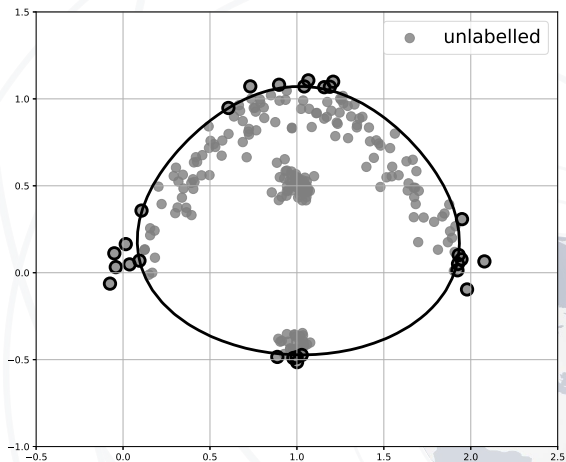
Proposition

- ▶ One-class SVM avec une partie des données labellisées

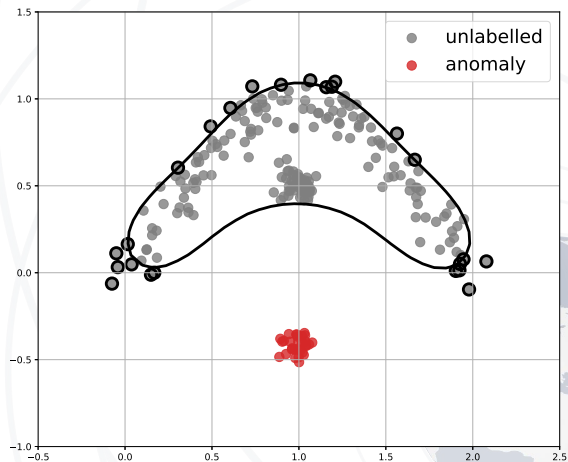
$$\begin{aligned} \arg \min_{\mathbf{w}, \rho, \xi} & \frac{1}{2} \|\mathbf{w}\|_2^2 - \rho + \frac{1}{\nu_1 n_1} \sum_{i=1}^{n_1} \xi_i + \frac{1}{\nu_2 n_2} \sum_{l=1}^{n_2} \zeta_l + \frac{1}{\nu_3 n_3} \sum_{p=1}^{n_3} \tau_p \\ \text{s.t. } & \mathbf{w}^T \Phi(\mathbf{x}_i) - \rho \geq -\xi_i \text{ and } \xi_i \geq 0 \quad \text{données non-labellisées} \\ & \mathbf{w}^T \Phi(\mathbf{y}_l) - \rho \geq -\zeta_l \text{ and } \zeta_l \geq 0 \quad \text{données normales} \\ & \mathbf{w}^T \Phi(\mathbf{z}_p) - \rho \leq \tau_p \text{ and } \tau_p \geq 0 \quad \text{anomalies} \end{aligned}$$

9. Julien LESOUPLE et Jean-Yves TOURNERET. "Incorporating User Feedback Into One-Class Support Vector Machines for Anomaly Detection". In : *Proc. Eur. Conf. on Signal Processing (EUSIPCO)*. Amsterdam, Netherlands, 2020, p. 1608–1612.

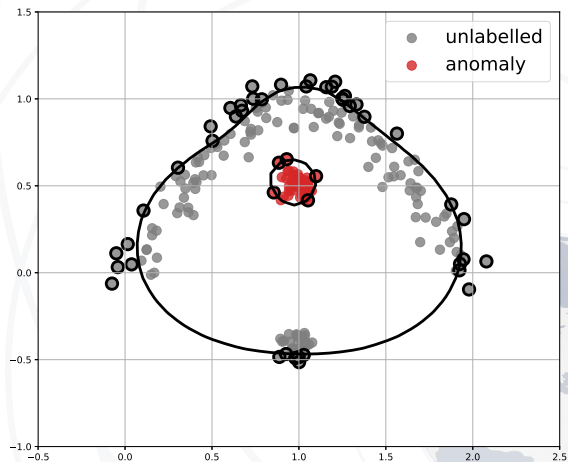
Illustration



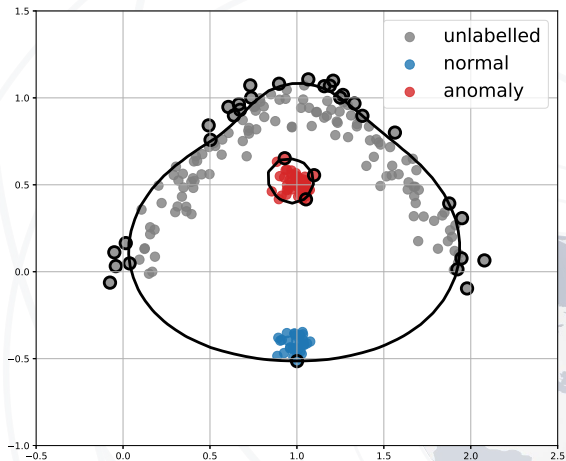
Illustration



Illustration



Illustration



$$\arg \min_{\mathbf{w}, \rho, \xi} \frac{1}{2} \|\mathbf{w}\|_2^2 - \rho + \frac{1}{\nu_1 n_1} \sum_{i=1}^{n_1} \xi_i + \frac{1}{\nu_2 n_2} \sum_{l=1}^{n_2} \zeta_l + \frac{1}{\nu_3 n_3} \sum_{p=1}^{n_3} \tau_p$$

s.t. $\mathbf{w}^T \Phi(\mathbf{x}_i) - \rho \geq -\xi_i$ et $\xi_i \geq 0$ données non-labellisées
 $\mathbf{w}^T \Phi(\mathbf{y}_l) - \rho \geq -\zeta_l$ et $\zeta_l \geq 0$ données normales
 $\mathbf{w}^T \Phi(\mathbf{z}_p) - \rho \leq \tau_p$ et $\tau_p \geq 0$ anomalies

Interprétation des hyperparamètres ?
Cas limites ?



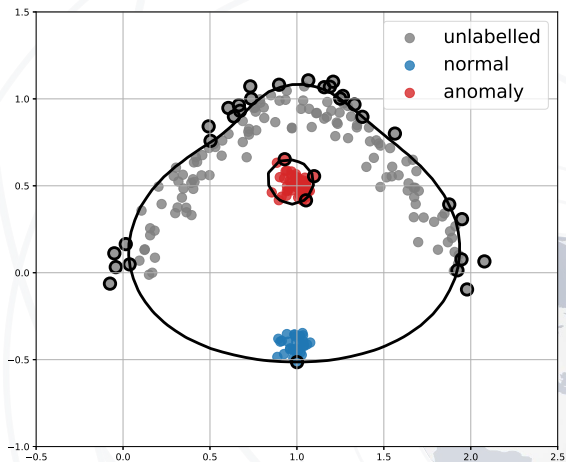
Reformulation : ν -SSVM¹⁰

$$\begin{aligned} \arg \min_{\mathbf{w} \in \mathbb{R}^M, \xi \in \mathbb{R}^n, b, \rho_1, \rho_2 \in \mathbb{R}} \quad & \frac{1}{2} \|\mathbf{w}\|_2^2 - r\rho_1 - (n-r)(\rho_2 - b) + \frac{1}{\nu_1} \sum_{i=1}^r \xi_i + \frac{1}{\nu_2} \sum_{i=r+1}^n \xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq \rho_1 - \xi_i, \quad \text{données labellisées} \\ & \mathbf{w}^T \Phi(\mathbf{x}_i) + b \geq \rho_2 - \xi_i, \quad \text{données non-labellisées} \\ & \xi_i \geq 0, \quad i = 1, \dots, n \\ & \rho_1 \geq 0, \rho_2 \geq 0 \end{aligned}$$

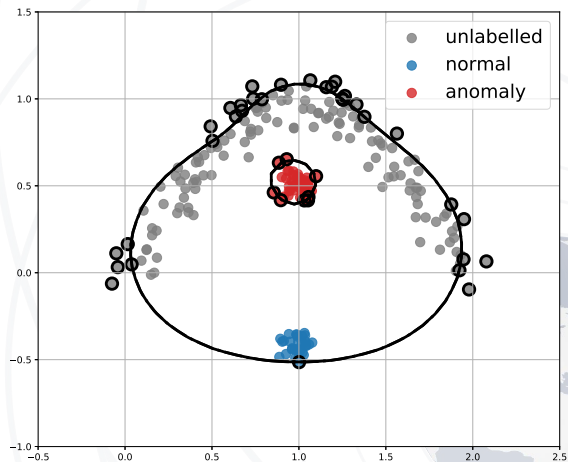
Hyperparamètres : proportion max de données mal classées ou du côté -1
Cas limites : nu-SVM et One Class-SVM

10. Julien LESOUPLE, Cédric BAUDOIN, Marc SPIGAI et Jean-Yves TURNERET. "Semi-Supervised Learning with Support Vector Machines for Anomaly Detection". In : submitted to *Signal Processing (under second review)* (2021).

Illustration



Illustration

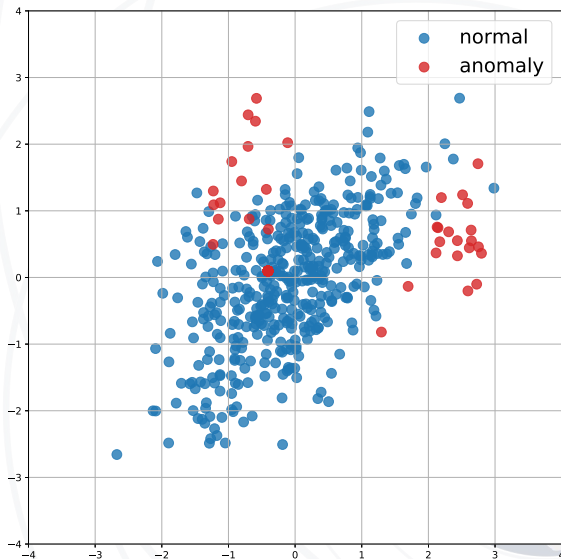


Algorithme itératif

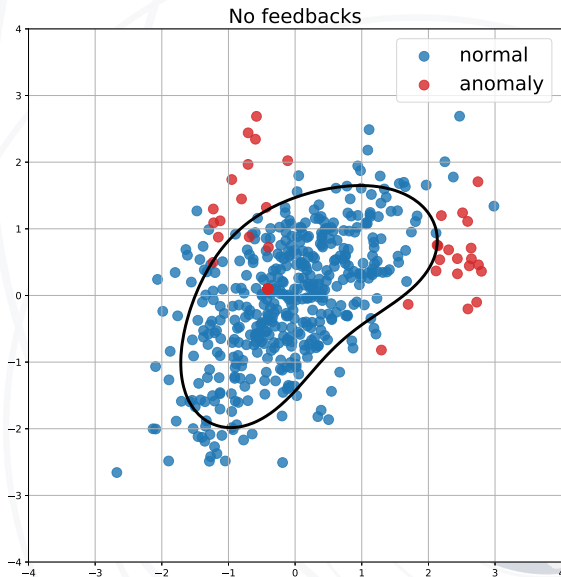
- ▶ initialisation par One Class SVM
- ▶ boucler jusqu'à un nombre max de requêtes à l'expert
 - ▶ calcul des scores d'anomalies pour les données non labellisées
 - ▶ demande d'un label pour la donnée avec le plus haut score
 - ▶ calcul de la nouvelle frontière de décision via ν -SSVM

11. Julien LESOUPLE, Cédric BAUDOUIN, Andrés TROYA-GALVIS et Jean-Yves TOURNERET. "Incorporating User Feedback in One-Class Support Vector Machines and Isolation Forest for Anomaly Detection". In : to be submitted to *Signal Processing* (2021).

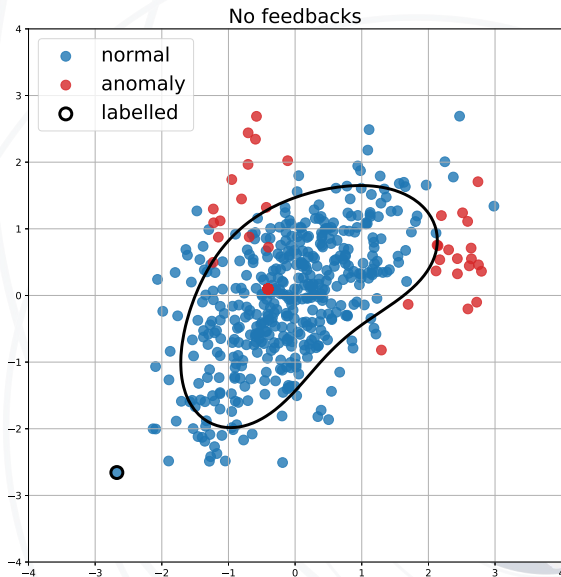
Illustration



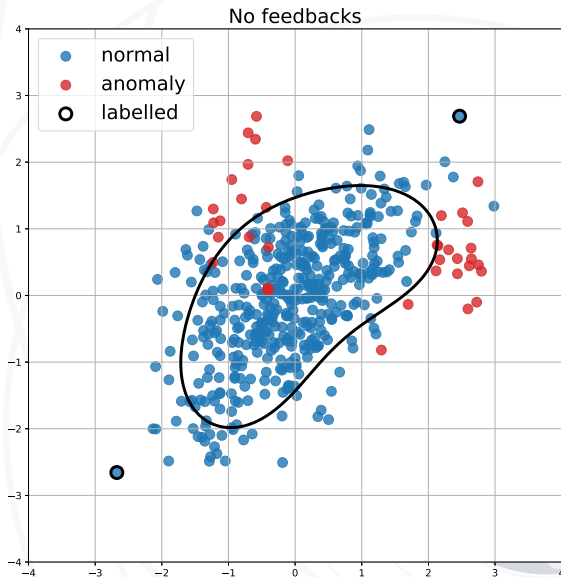
Illustration



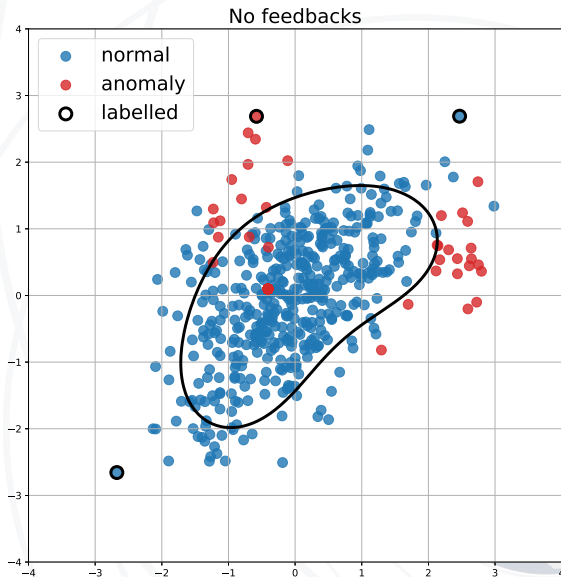
Illustration



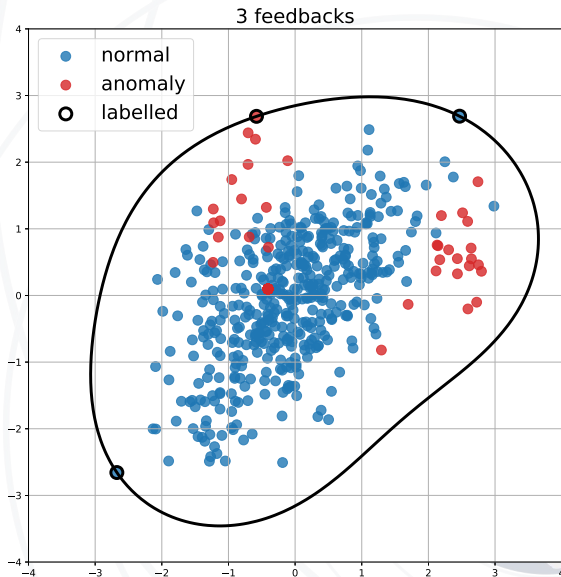
Illustration



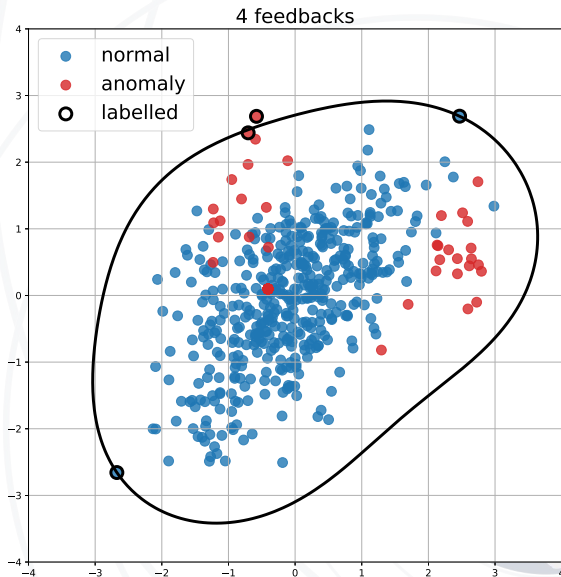
Illustration



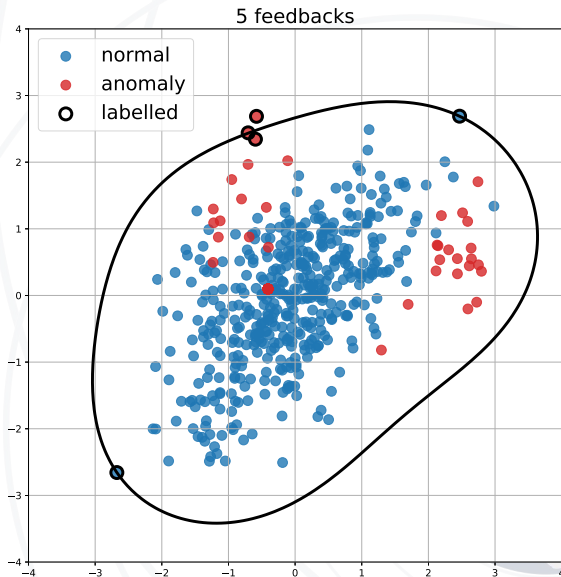
Illustration



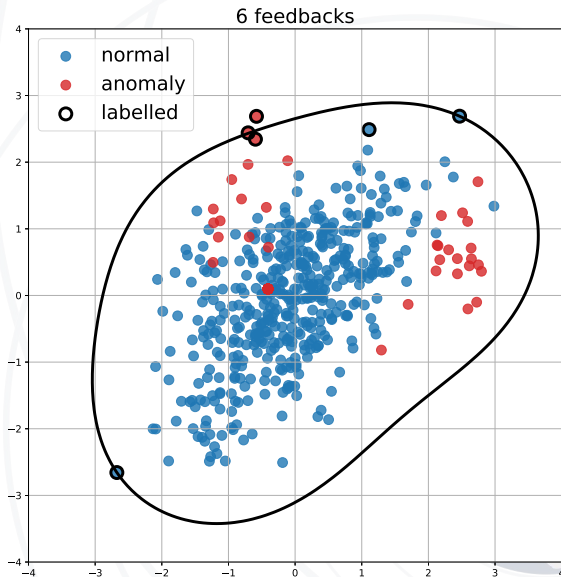
Illustration



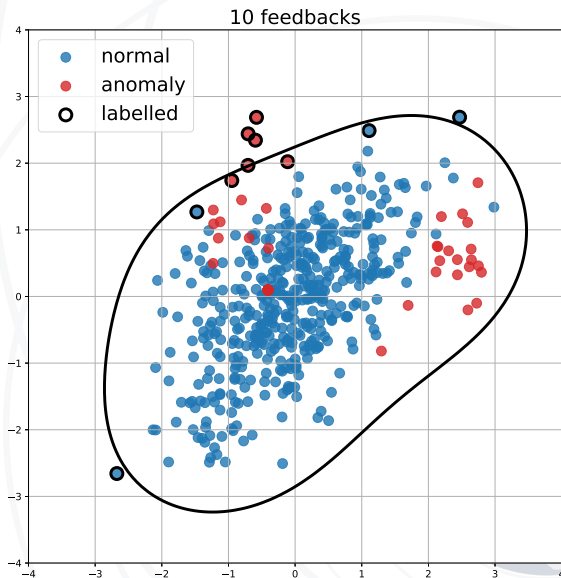
Illustration



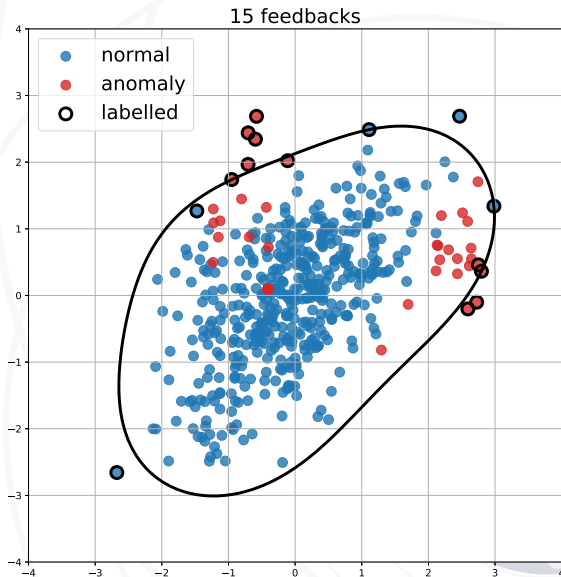
Illustration



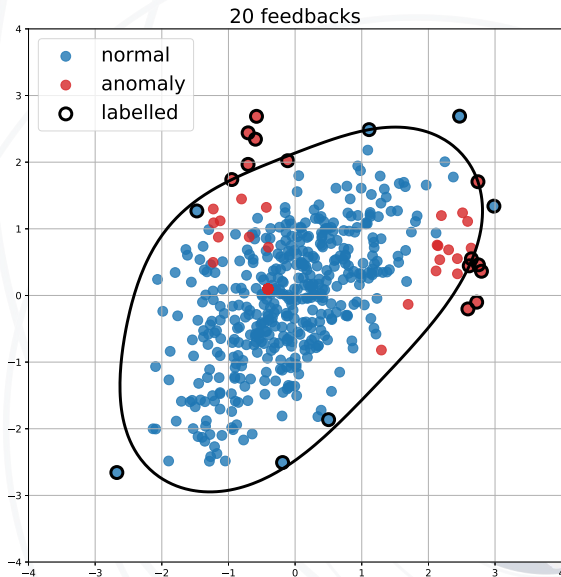
Illustration



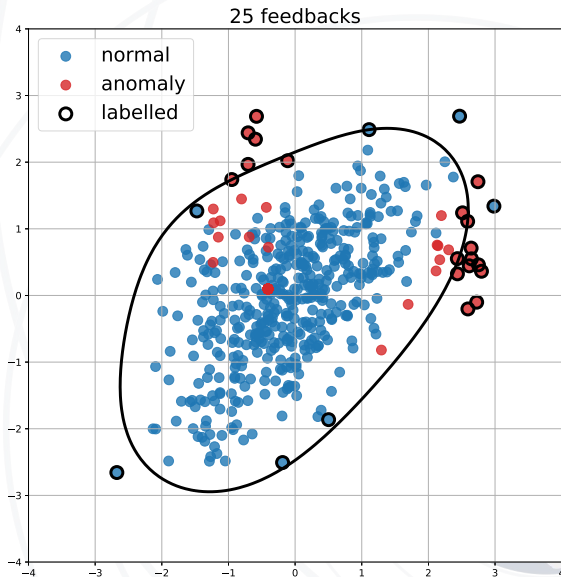
Illustration



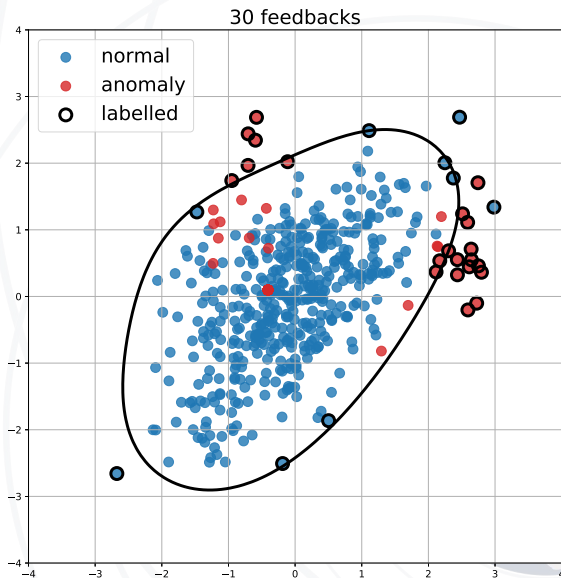
Illustration



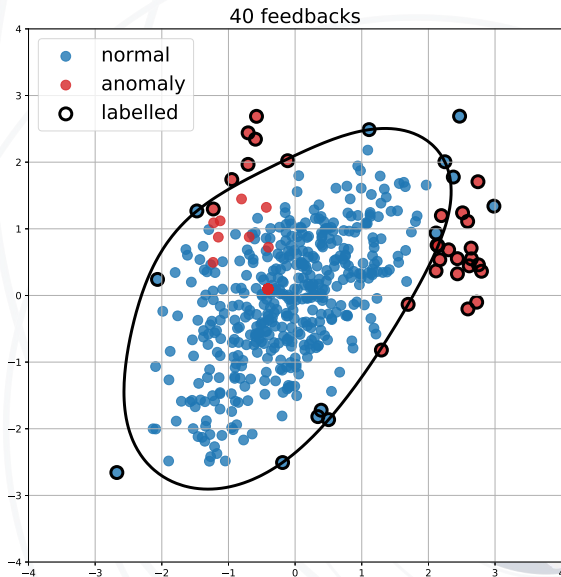
Illustration



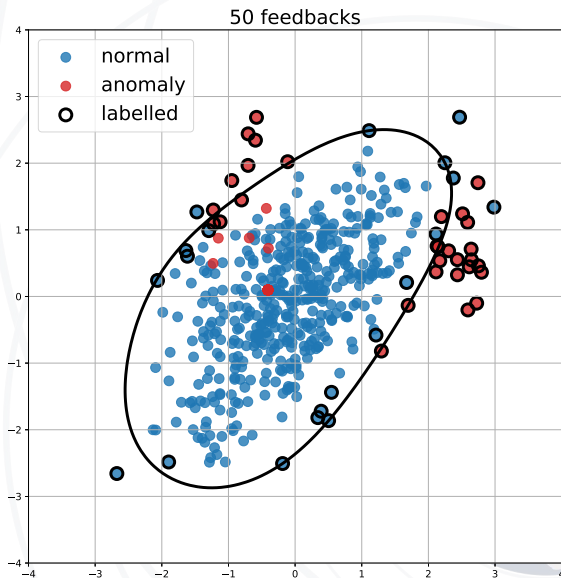
Illustration



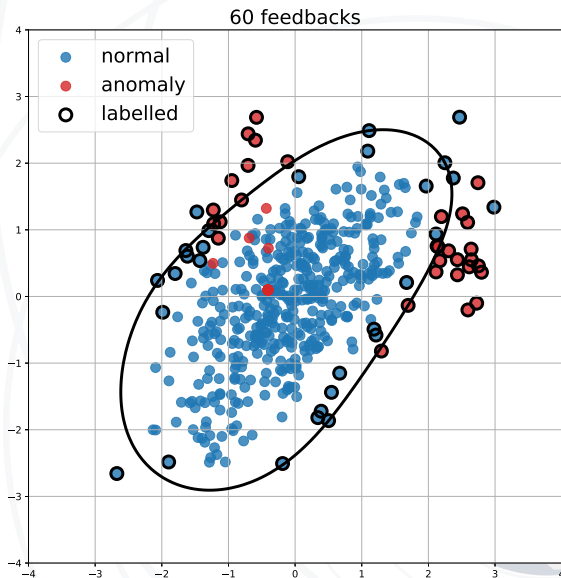
Illustration



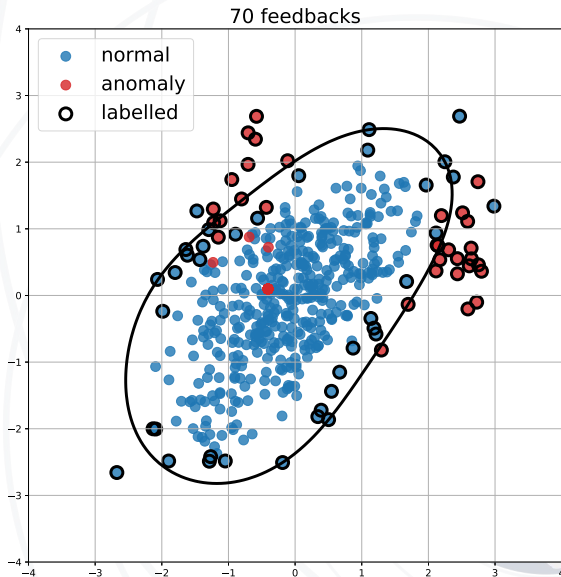
Illustration



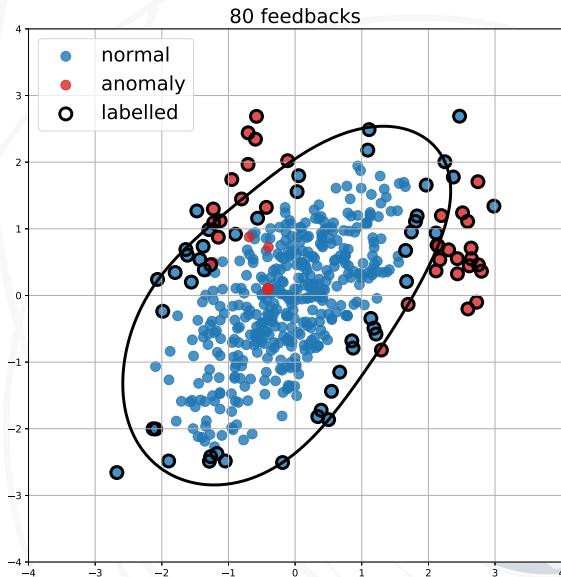
Illustration



Illustration



Illustration



Plan

Contexte

Machines à vecteurs supports

Détection d'anomalies

Ajout de retour utilisateur

Conclusion



Conclusion (1/2)

Dans ce séminaire

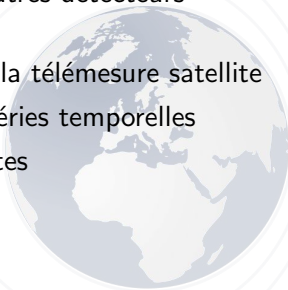
- ▶ présentation des machines à vecteurs supports
- ▶ application à la détection d'anomalies avec retour utilisateur

Dans nos travaux actuels

- ▶ application du retour utilisateur à d'autres détecteurs d'anomalies (notamment IF)

Futurs travaux : appliquer ces méthodes à la télémessure satellite

- ▶ application de ces algorithmes à des séries temporelles
- ▶ prise en compte de données manquantes



Conclusion (2/2)

Merci pour votre attention ! Des questions ?

Contact : `julien.lesouple@tesa.prd.fr`

Codes disponibles : `http://perso.tesa.prd.fr/jlesouple`

