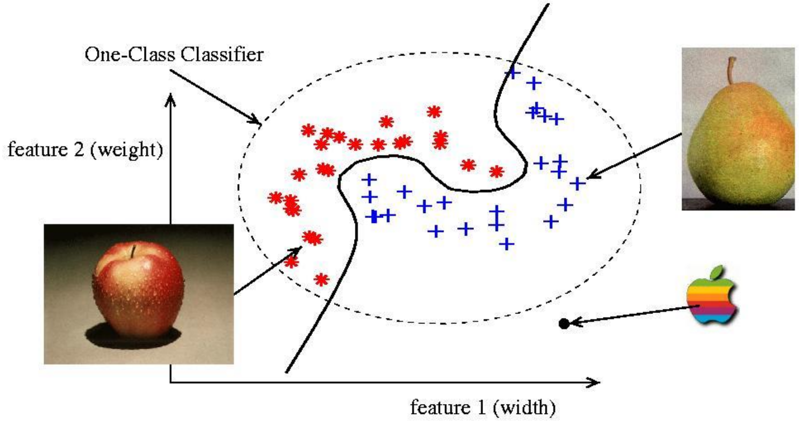
Detecção de Novidades

André E. Lazzaretti

UTFPR/CPGEI

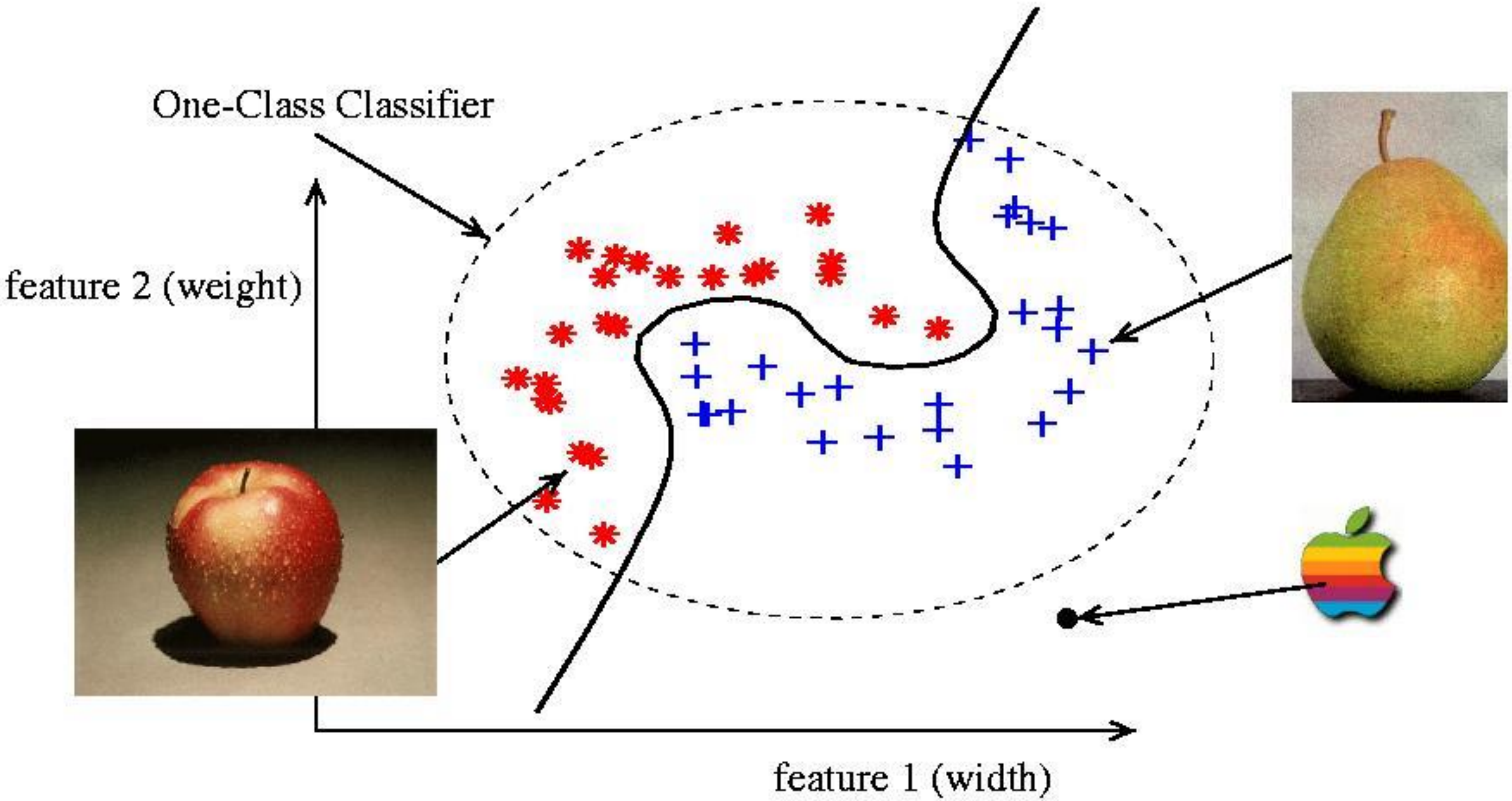


UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E INFORMÁTICA INDUSTRIAL

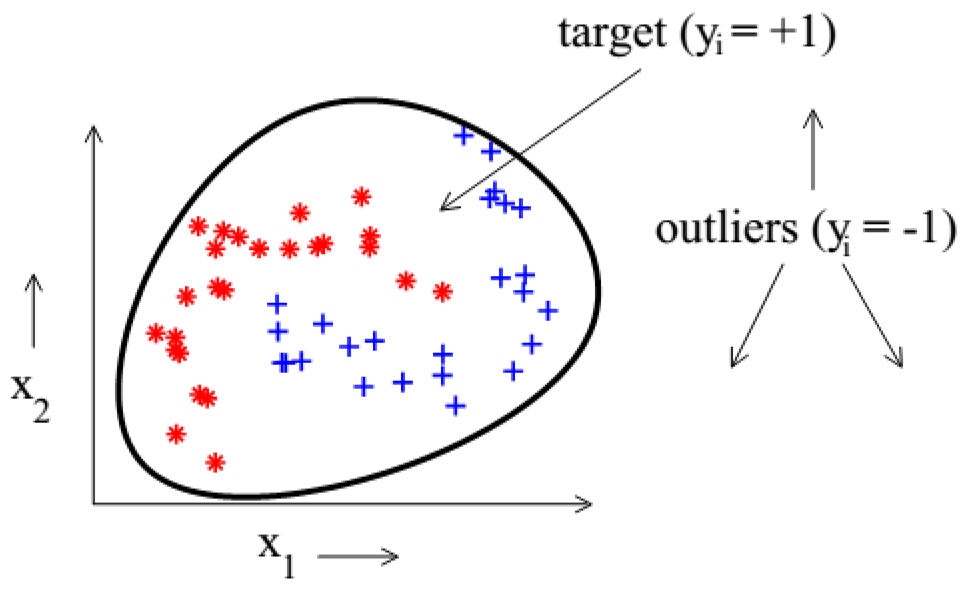
Detecção de Novidades

- Ideia Geral:



Fundamentação

- **Sinônimos**: outlier detection, anomaly detection, one-class classification.
- **Formulação**: somente targets (normais) disponíveis durante o treinamento:

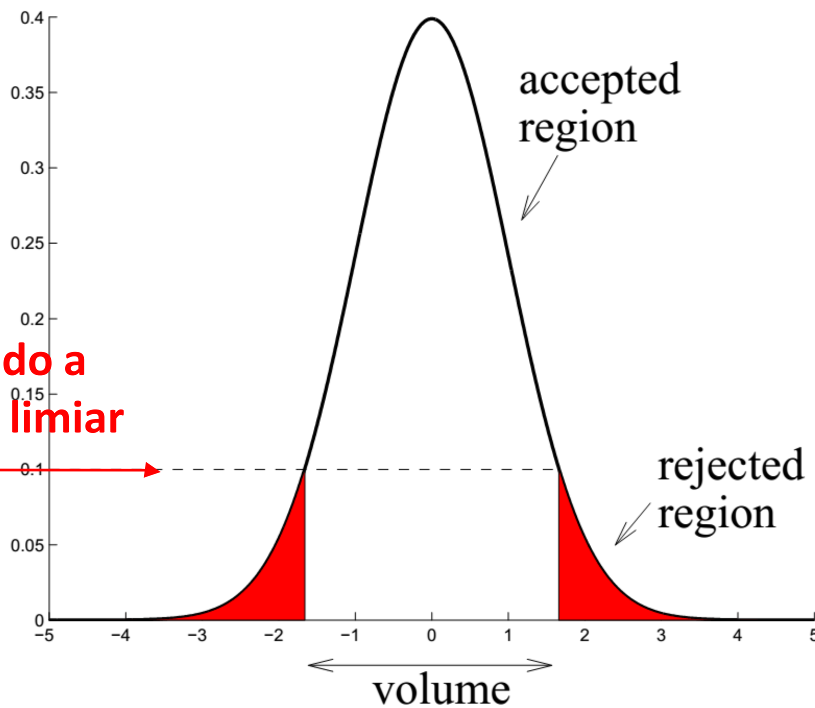


Principais Métodos

- *Density-Estimation Methods*:
  - Paramétricos (Gaussianas) e Não-Paramétricos (Parzen)
- *Boundaries Methods*:
  - Foca apenas na superfície de decisão ao invés de modelar a distribuição “completa”
- *Reconstruction Methods*:
  - Erro de reconstrução (autoencoder)
  - Agrupamento (clustering)

Density Methods

- Assumindo distribuição Normal:



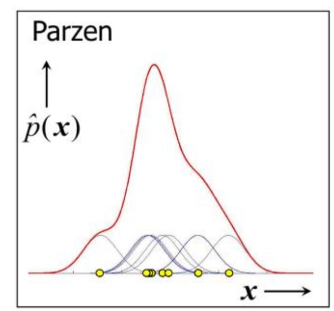
Comparado a um dado limiar

$$p_{\mathcal{N}}(\mathbf{z};\boldsymbol{\mu},\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\}$$

Também:

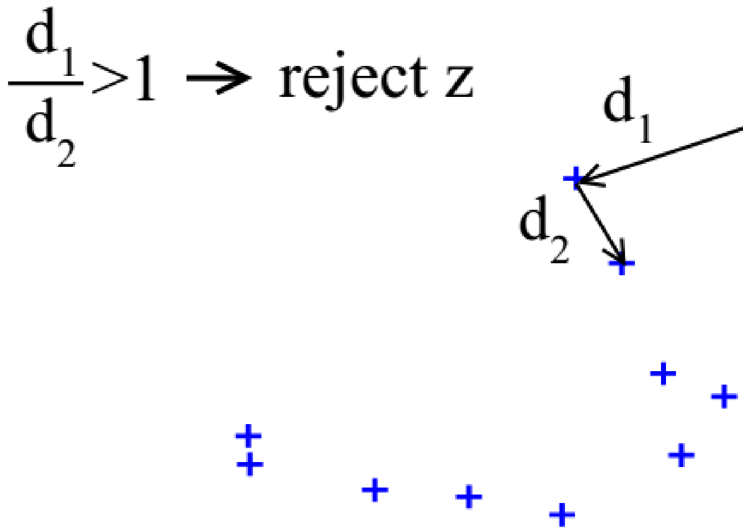
$$p_{MoG}(\mathbf{x}) = \frac{1}{N_{MoG}}\sum_j \alpha_j p_{\mathcal{N}}(\mathbf{x};\boldsymbol{\mu}_j,\Sigma_j)$$

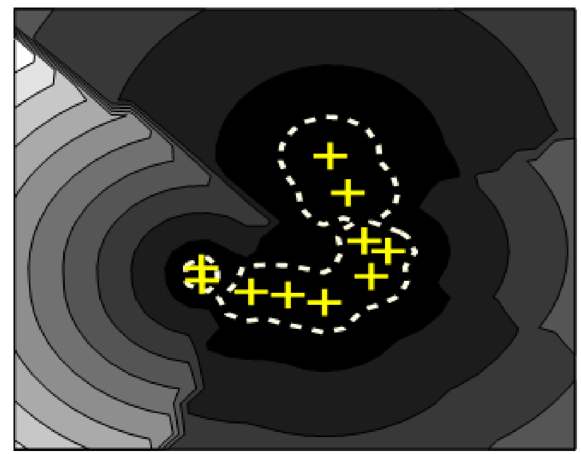
Ou:



Vizinho-mais-próximo

- Visualmente:





$$f_{NN^{tr}}(\mathbf{z}) = I\left(\frac{\|\mathbf{z} - NN^{tr}(\mathbf{z})\|}{\|NN^{tr}(\mathbf{z}) - NN^{tr}(NN^{tr}(\mathbf{z}))\|} \leq 1\right)$$

Sendo:

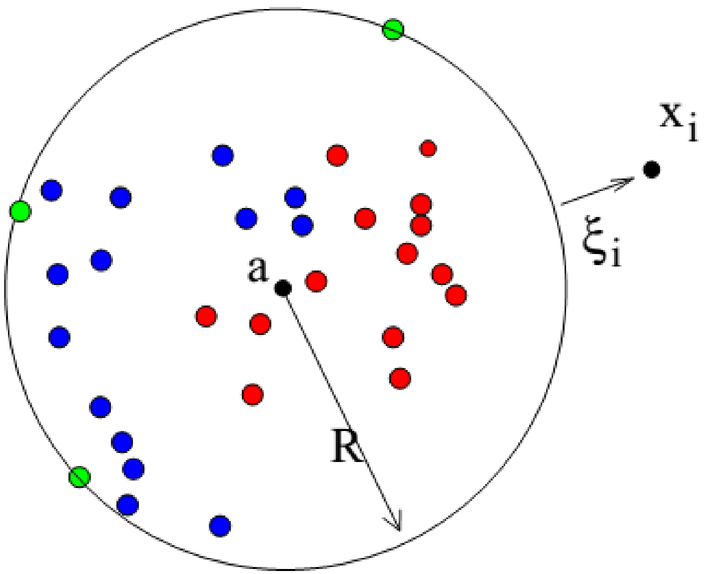
$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true,} \\ 0 & \text{otherwise.} \end{cases}$$

Variantes:

- Usar a média dos k vizinhos mais próximos;
- Ao invés de utilizar "1", incluir percentual;
- Alterar a medida de distância.

Support Vector Data Description

- Modelagem por uma hiperesfera:



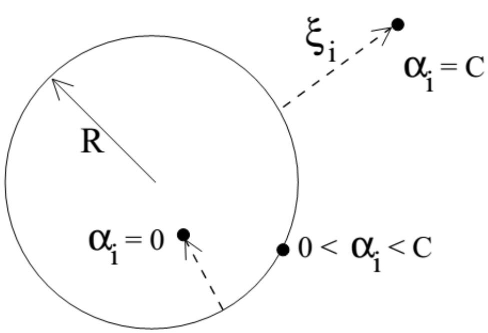
$$\mathcal{E}(R, \mathbf{a}, \boldsymbol{\xi}) = R^2 + C \sum_i \xi_i$$

$$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \xi_i, \qquad \xi_i \geq 0, \qquad \forall i$$

Representação de Wolfe (Exercício!)

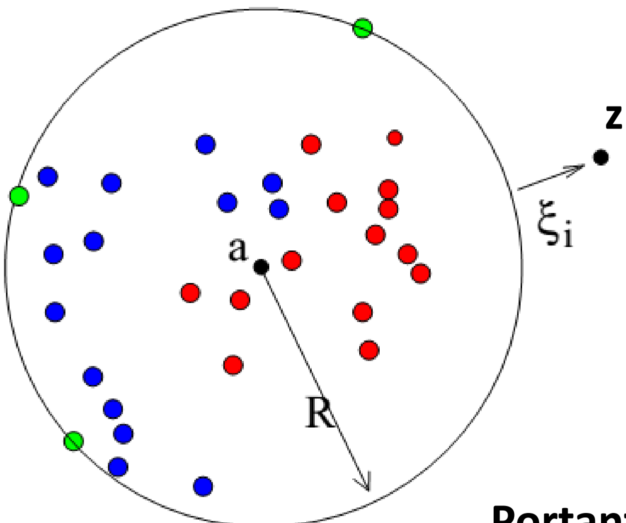
$$L = \sum_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}_i) - \sum_{i,j} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$0 \leq \alpha_i \leq C$$



Support Vector Data Description

- Classificação de um novo exemplo **z**:



$$\|\mathbf{z}-\mathbf{a}\|^2 = (\mathbf{z}\cdot\mathbf{z}) - 2\sum_i \alpha_i(\mathbf{z}\cdot\mathbf{x}_i) + \sum_{i,j}\alpha_i\alpha_j(\mathbf{x}_i\cdot\mathbf{x}_j) \le R^2$$

Sendo: $$\mathbf{a} = \sum_i \alpha_i \mathbf{x}_i$$

$$R^2 = (\mathbf{x}_k\cdot\mathbf{x}_k) - 2\sum_i \alpha_i(\mathbf{x}_i\cdot\mathbf{x}_k) + \sum_{i,j}\alpha_i\alpha_j(\mathbf{x}_i\cdot\mathbf{x}_j)$$

Portanto:

$$\begin{aligned} f_{SVDD}(\mathbf{z};\boldsymbol{\alpha},R) &= I\left(\|\mathbf{z}-\mathbf{a}\|^2 \le R^2\right) \\ &= I\left((\mathbf{z}\cdot\mathbf{z}) - 2\sum_i \alpha_i(\mathbf{z}\cdot\mathbf{x}_i) + \sum_{i,j}\alpha_i\alpha_j(\mathbf{x}_i\cdot\mathbf{x}_j) \le R^2\right) \end{aligned}$$

Sendo:

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true,} \\ 0 & \text{otherwise.} \end{cases}$$

Support Vector Data Description with Negative Examples

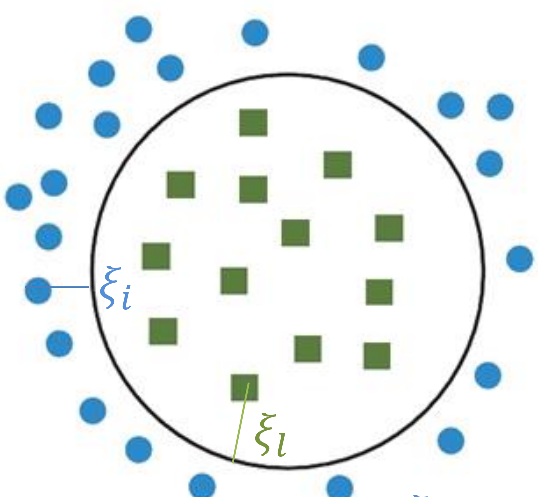
- Supondo que uma determinada quantidade de outliers (negative examples) estão disponíveis:

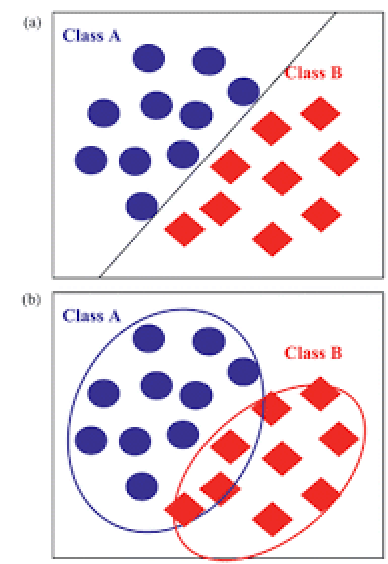
$$\mathcal{E}(R, \mathbf{a}, \boldsymbol{\xi}) = R^2 + C_1 \sum_i \xi_i + C_2 \sum_l \xi_l$$

$$\|\mathbf{x}_i - \mathbf{a}\|^2 \le R^2 + \xi_i,$$

$$\|\mathbf{x}_l - \mathbf{a}\|^2 > R^2 - \xi_l,$$

$$\xi_i \ge 0, \xi_l \ge 0 \quad \forall i, l$$





ξ_i *– positive muito distante*

ξ_l *– negative muito próximo*

Descrições Mais Flexíveis

- Supondo um mapeamento: $\mathbf{x}^* = \Phi(\mathbf{x})$

$$L = \sum_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}_i) - \sum_{i,j} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\Downarrow$$

$$L = \sum_i \alpha_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_i) - \sum_{i,j} \alpha_i \alpha_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$$

$$\Downarrow$$

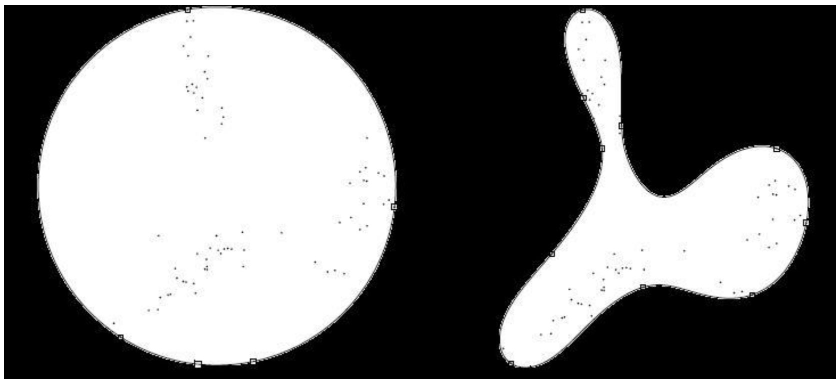
$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$$

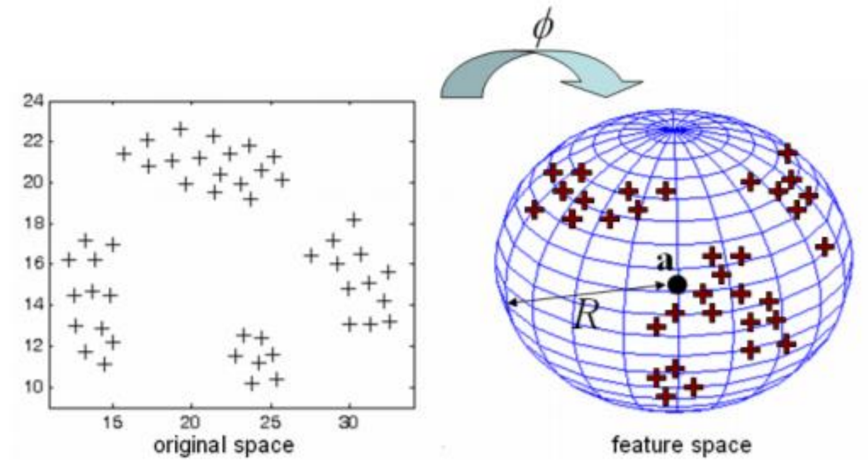
$$\Downarrow$$

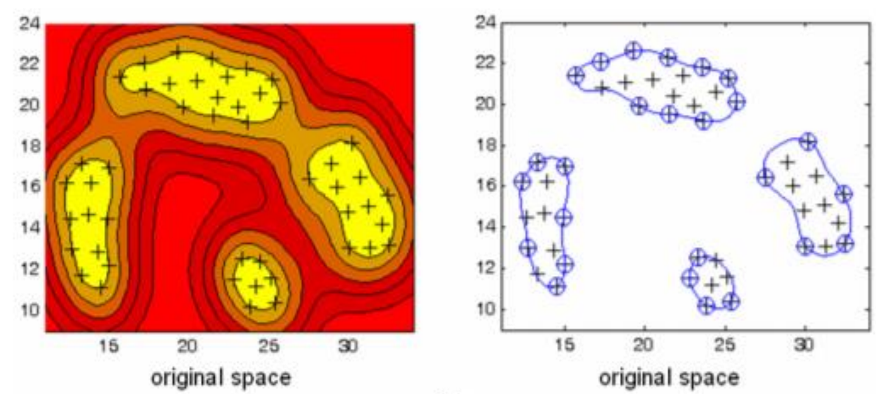
$$L = \sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$$

Descrições Mais Flexíveis

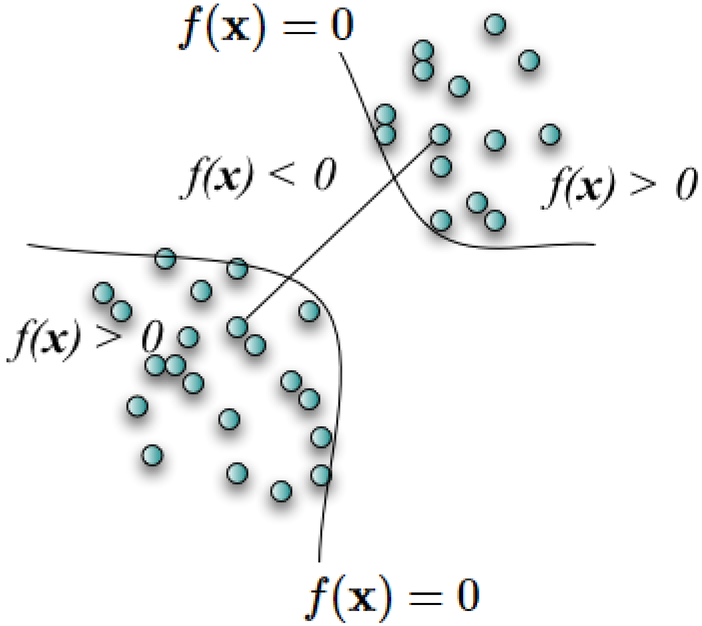
- Com isso:

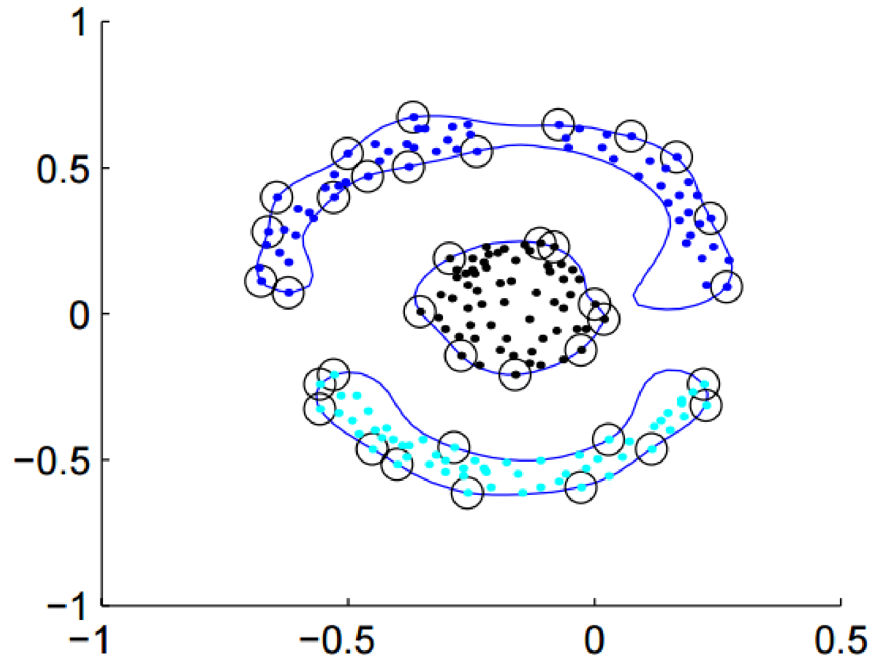






Support Vector Clustering





Matriz de Relação:

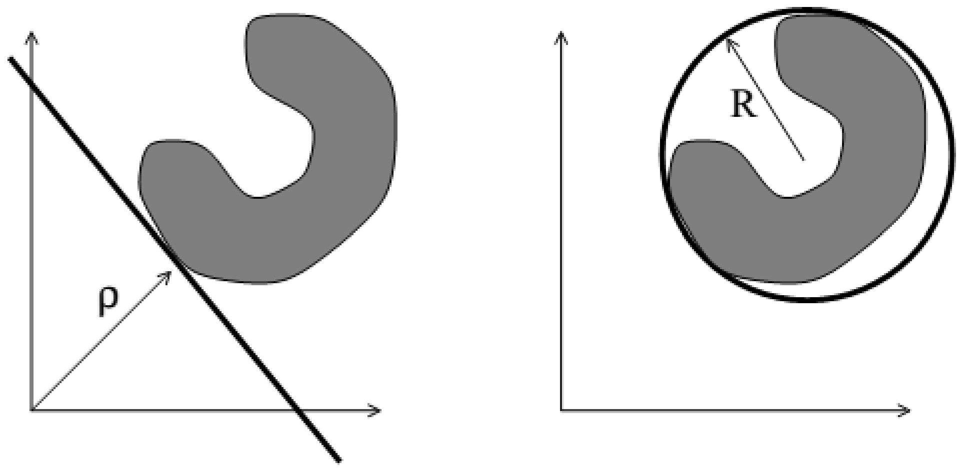
$$A_{ij} = \begin{cases} 1, & \text{if } f(\mathbf{x}) > 0 \text{ for all } \mathbf{x} \text{ on the line segment connecting } \mathbf{x_i} \text{ and } \mathbf{x_j} \\ 0 & \text{otherwise.} \end{cases}$$

Clusters are defined as the connected components of the graph induced by **A**

Relação com ν-SVM

- SVDD x One-Class SVM (ν-SVM):

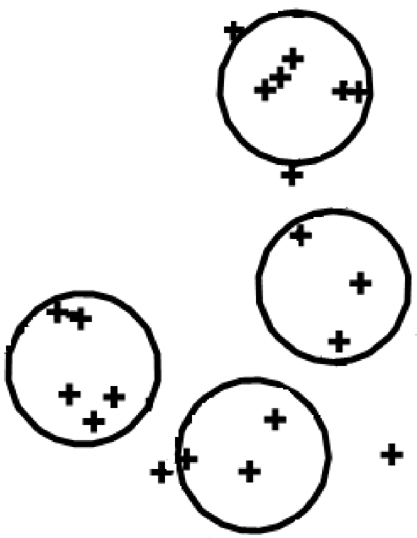
$$\mathcal{E}(R, \mathbf{a}, \boldsymbol{\xi}) = R^2 + C \sum_i \xi_i$$



$$\min \left(\frac{1}{2} \|\mathbf{w}\|^2 - \rho + \frac{1}{\nu N} \sum_i \xi_i \right)$$

Reconstruction Methods

- K-Means:
  - Aplica-se o agrupamento;
  - Na etapa de teste, a distância mínima do exemplo de teste e o centro mais próximo pode ser utilizada para quantificar a anormalidade.



Teste: Efetua-se o cálculo das distâncias entre cada padrão e os centros dos agrupamentos mais próximos e, em seguida, utiliza-se a média dos N_R valores mais distantes para calcular o limiar e definir, posteriormente, se um novo padrão é novidade ou faz parte do conjunto de dados normais. N_R deve ser definido a priori.

Exemplos MATLAB!

Semi-Supervised Learning

- **Sinônimos**: *transductive (inductive) learning*;
- **Cenário**: pequena quantidade de dados rotulados e uma enorme quantidade de dados não rotulados;
- **Ideia**: incluir os dados não rotulados no processo de treinamento, de forma que a superfície de decisão esteja localizada em uma região com poucos dados (baixa densidade):

Semi-Supervised Learning

- Premissas:
  - Exemplos que estão próximos no espaço de características tendem a ter o mesmo rótulo.
  - Os dados tendem a formar clusters e exemplos em um mesmo cluster tendem a ter o mesmo rótulo, muito embora dados com os mesmos rótulos podem estar espalhados em múltiplos clusters.
  - Os dados tendem a formar um *manifold* de dimensão inferior à dimensão do espaço de entrada.

Semi-Supervised SVMs (S3VM ou TSVM)

- Formulação (hard-margin), l – label. u – unlab.:

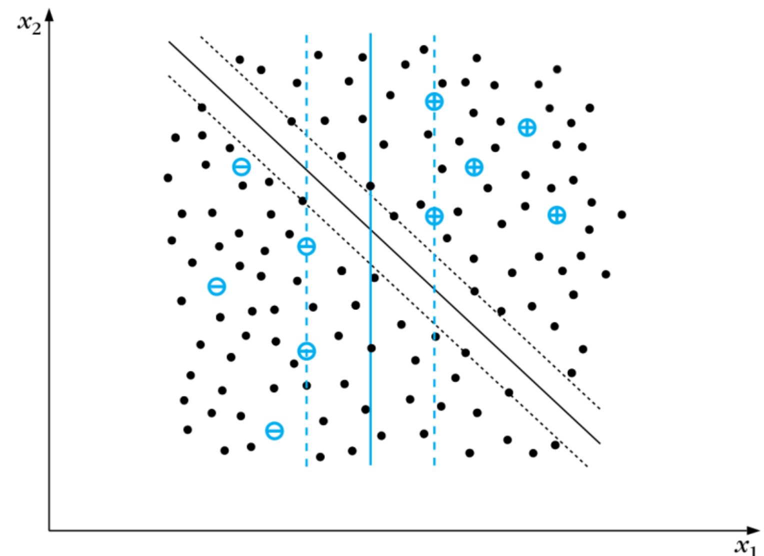
$$\text{minimize} \quad J(y_{N_l+1}, \ldots, y_{N_l+N_u}, \boldsymbol{w}, w_0) = \frac{1}{2}||\boldsymbol{w}||^2$$

$$\text{subject to} \quad y_i(\boldsymbol{w}^T \boldsymbol{x}_i + w_0) \geq 1, \; i = 1, 2, \ldots, N_l$$

$$y_i(\boldsymbol{w}^T \boldsymbol{x}_i + w_0) \geq 1, \; i = N_l + 1, \ldots, N_l + N_u$$

$$\boxed{y_i \in \{+1, -1\}, \; i = N_l + 1, \ldots, N_l + N_u}$$

São variáveis do problema de otimização



- Problema: y_i, $i=N_l+1, \ldots, N_l + N_u$, são valores inteiros em {+1, -1} – otimização não-convexa.

- Existem alternativas para aproximar, mas ainda é um problema em aberto.

Referências

- Tese Dr. David M. J. Tax
- Artigo Ben Hur – Support Vector Clustering