

**Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Pós-Graduação em Ciência da Computação**

Inteligência Artificial

Aprendizagem Outras Técnicas

Prof.^a Joseana Macêdo Fachine Régis de Araújo
joseana@computacao.ufcg.edu.br

Aprendizagem

Tópico

- Aprendizagem
 - Outras Técnicas



Aprendizagem de Máquina

Abordagens

- Simbólica
- Conexionista
- Analítica
- Evolutiva
- Estatística

- **Problemas**
- Classificação, Agrupamento, ...

Aprendizagem de Máquina

Técnicas de Aprendizagem

- ❑ Aprendizagem Bayesiana
- ❑ Algoritmos Genéticos
- ❑ Redes Neurais Artificiais
- ❑ *Deep Learning*
- ❑ Árvores de Decisão
- ❑ Agrupamento (K-Means)
- ❑ Máquinas de Vetores de Suporte (SVM), ...

Supervisionada
Não Supervisionada
Por Reforço

IMPORTANTE:
vetor de
características

Agrupamento

O QUE É AGRUPAMENTO DE DADOS?

- ▣ A organização de dados em grupos é uma das formas mais fundamentais para entendimento e aprendizado.
- ▣ Análise de grupos ou clusters é o estudo de algoritmos e métodos para agrupar objetos de acordo com suas características.
- ▣ Dada uma representação de n objetos, encontrar K grupos baseando-se em uma medida de similaridade, tal que objetos dentro de um mesmo grupo são semelhantes e objetos de grupos diferentes são distintos.

Agrupamento

O que é similaridade?

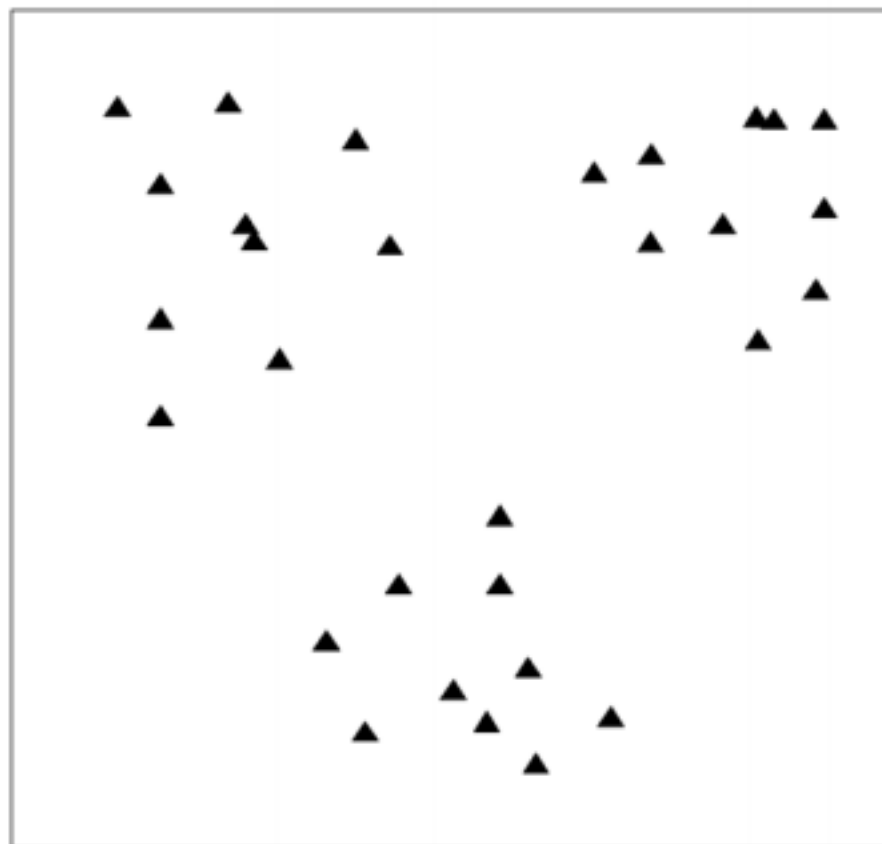
- Quão semelhantes são dois objetos diferentes de um conjunto de dados

O que é um grupo?

- Ideal: conjunto compacto e isolado de pontos (objetos)
- Real: subjetivo, nos olhos de quem vê

- Análise de grupos é utilizada, especialmente, em casos de dados multivariados.

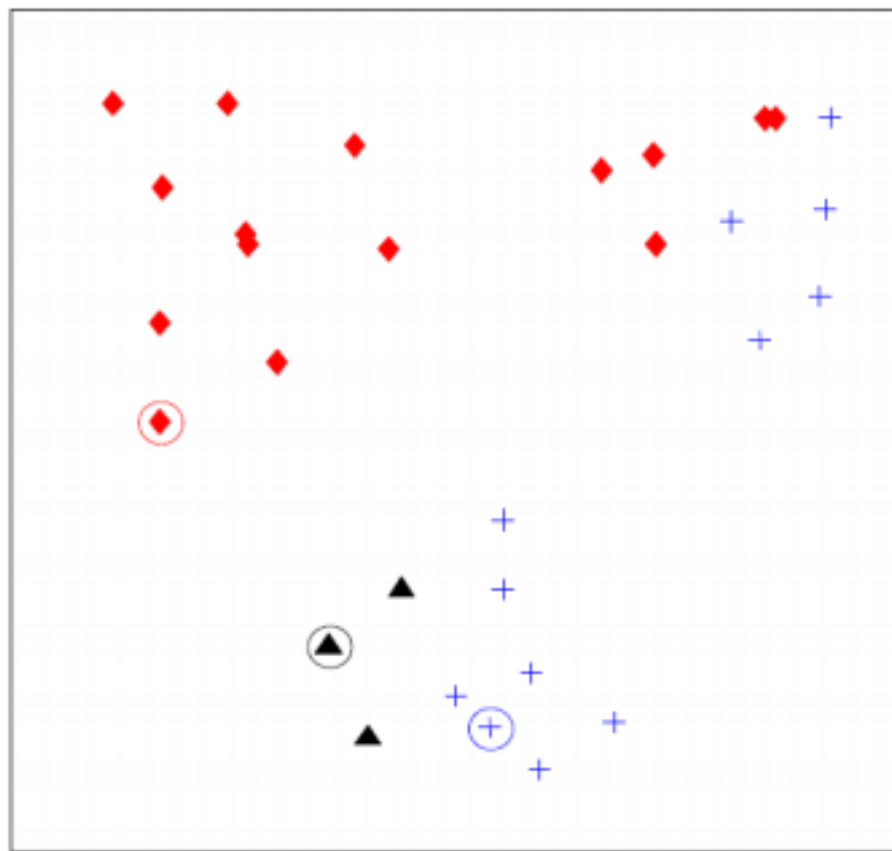
Agrupamento (K-means)



Conjunto X

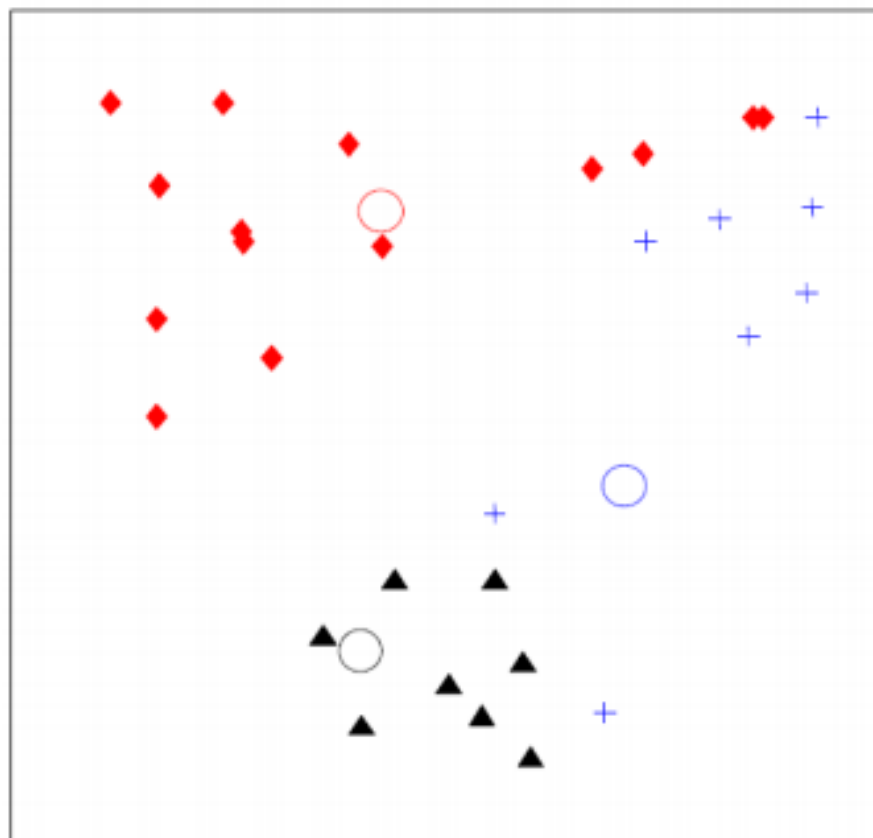
Fonte: http://www2.dc.ufscar.br/~priscilla_lopes/jornada/jornada.pdf

Agrupamento (K-means)



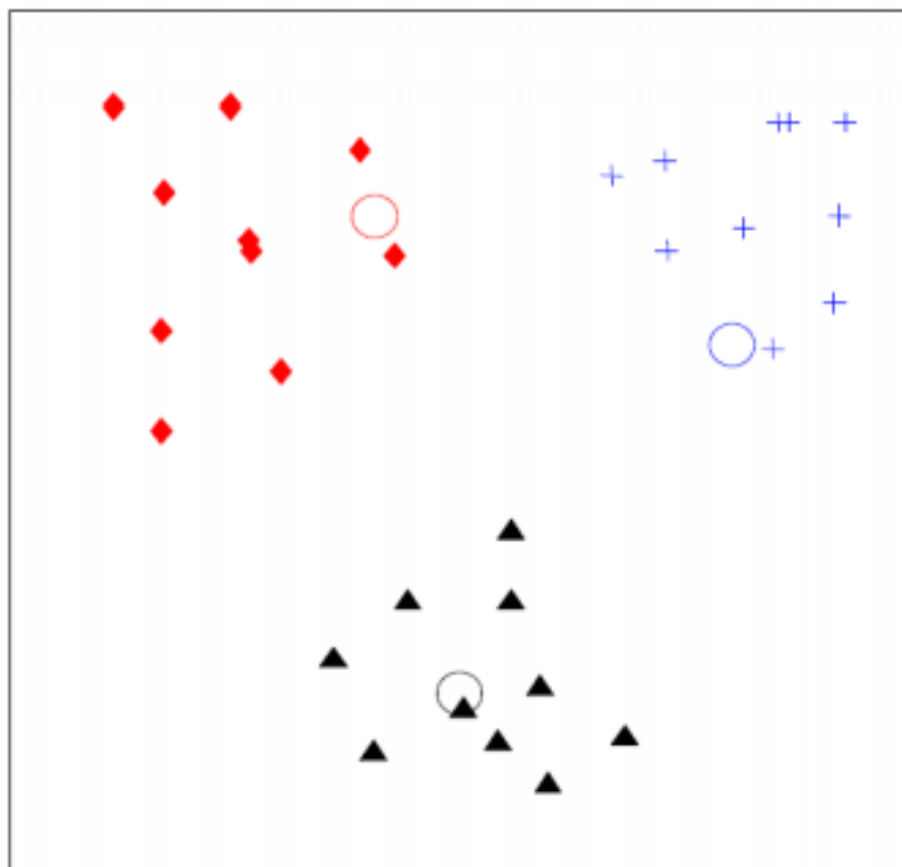
Definição dos centros

Agrupamento (K-means)



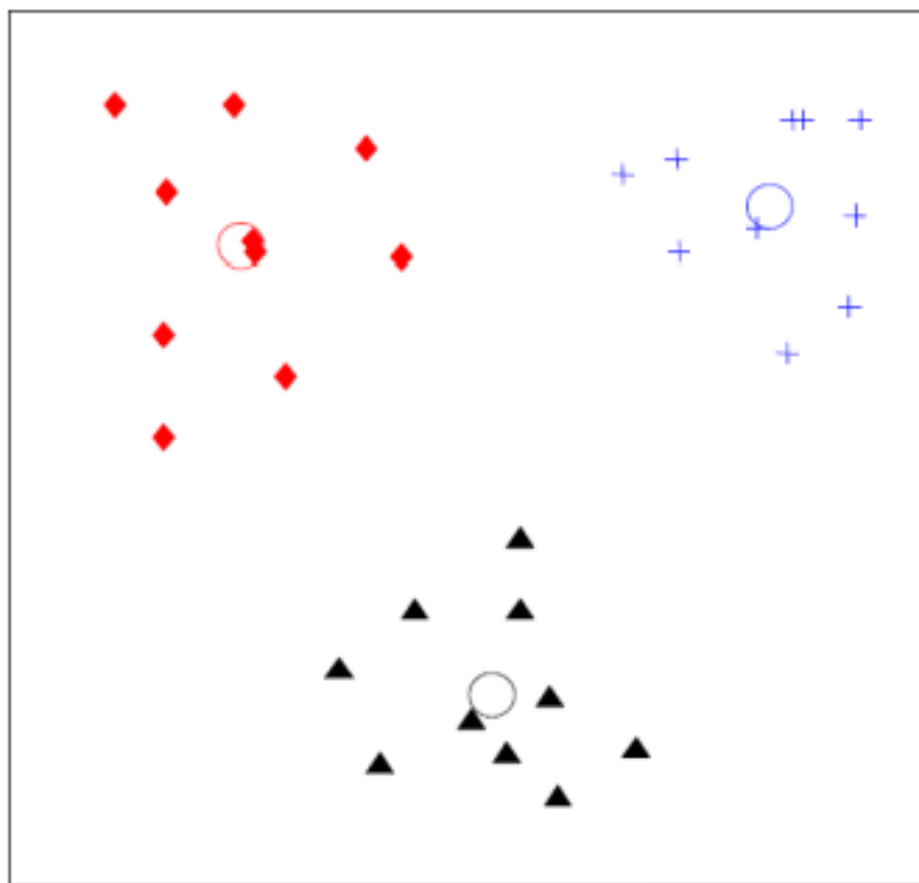
Iteração 2

Agrupamento (K-means)



Iteração 3

Agrupamento (K-means)



Final

Agrupamento (K-means)

- Algoritmo de aprendizado mais simples
- Este algoritmo supõe que todos os padrões (instâncias) são pontos no espaço n -dimensional R^n
- Os vizinhos mais próximos de um padrão são definidos em termos da distância Euclidiana padrão
 - Seja um padrão \mathbf{x} arbitrário descrito pelo vetor de características
 - $\langle a_1(\mathbf{x}), a_2(\mathbf{x}), \dots, a_n(\mathbf{x}) \rangle$, em que $a_r(\mathbf{x})$ representa o valor do r -ésimo atributo de \mathbf{x} , então a distância euclidiana entre \mathbf{x}_i e \mathbf{x}_j

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Agrupamento (K-means)

- Algoritmo de Treinamento:
 - Para cada padrão de treinamento $\langle \mathbf{x}, f(\mathbf{x}) \rangle$, adicione o exemplo a lista de *exemplos_de_treinamento*
- Algoritmo de Classificação:
 - Dado um padrão (instância) de consulta \mathbf{x}_q a ser classificado
 - Seja $\mathbf{x}_1, \dots, \mathbf{x}_k$ as k instâncias (padrões) do *exemplos_de_treinamento* que são mais próximos a \mathbf{x}_q . Retorne
- Um refinamento óbvio do k -NN é ponderar a contribuição de cada dos k vizinhos de acordo com a distância do ponto \mathbf{x}_q , dando maior peso para os vizinhos mais próximos



Agrupamento

Aplicações

- Segmentação de imagens
- Agrupamento de documentos
- Agrupamento de clientes para marketing eficiente
- Gerenciamento de equipe
- Estudo de dados de genoma

Support Vector Machines - SVM

Método supervisionado de aprendizagem para problemas de:

- Classificação
- Regressão

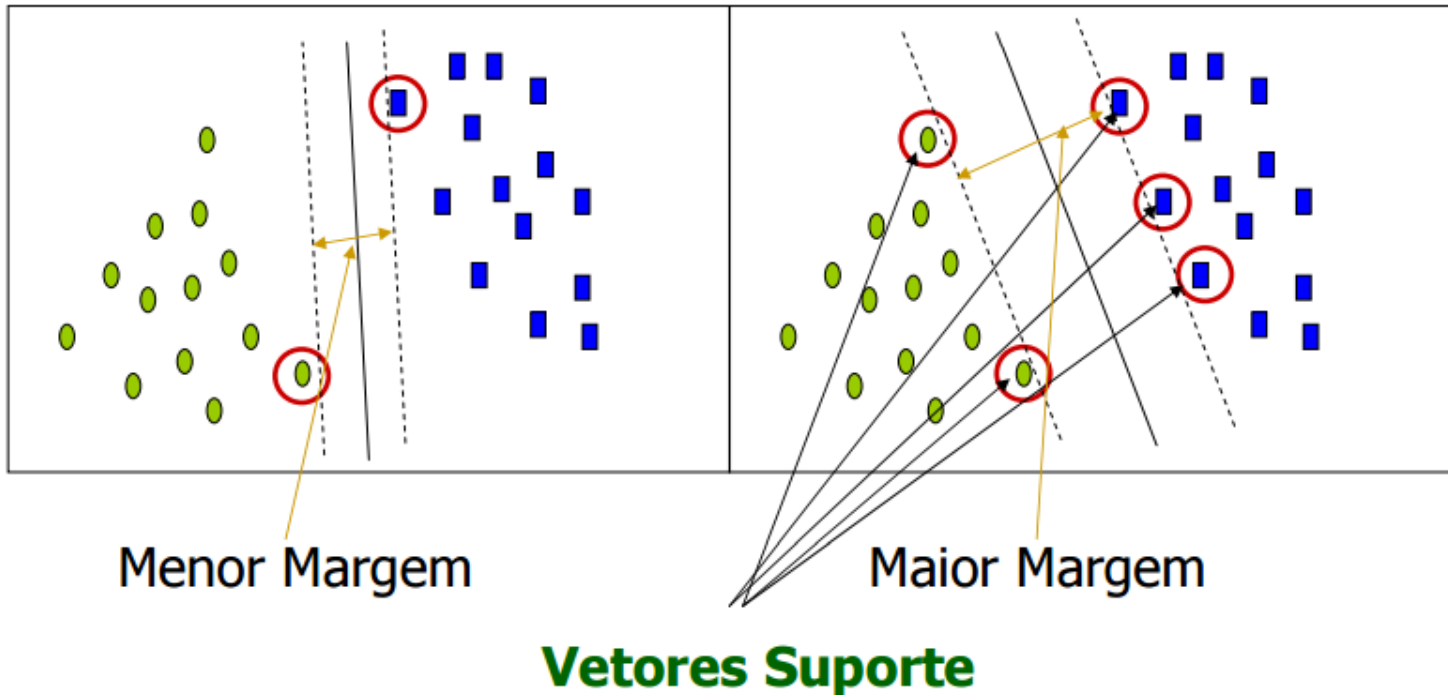
Duas ideias principais

- Assumindo classes linearmente separáveis, aprende hiperplanos de separação com margem máxima.
- Expande a entrada para um espaço de alta dimensionalidade para lidar com casos não-linearmente separáveis.

Support Vector Machines - SVM

- Método de classificação para dados lineares e não-lineares.
- SVM usa um mapeamento não-linear para transformar os dados originais (treinamento) em uma dimensão maior.
- Na nova dimensão, SVM busca um hiperplano que gere uma separação linear ótima.
- SVM sempre separa duas classes por meio de um hiperplano com a maior distância entre as classes.
- SVM encontra o hiperplano usando vetores suportes (tuplas especiais no treinamento) e margens definidas pelos vetores suportes.

Support Vector Machines - SVM

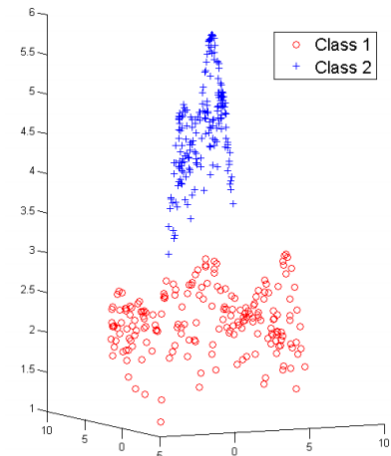
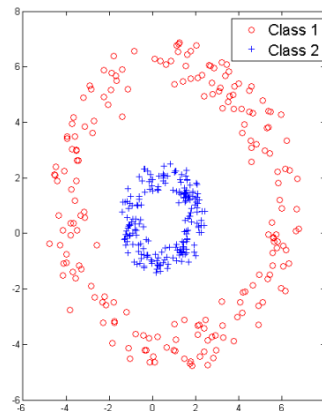
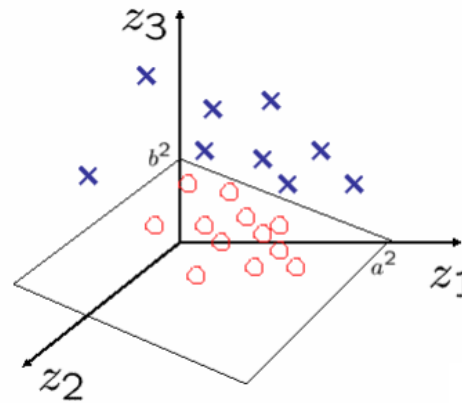
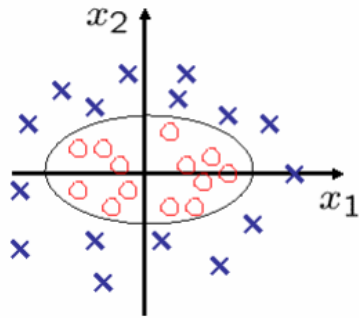


Os exemplos de treinamento que estiverem mais próximos da função de separação são chamados de vetores de suporte.

Fonte: <http://www.ime.unicamp.br/~wanderson/Aulas/MT803-Aula13-Bayes-RNA-SVM.pdf>

Support Vector Machines - SVM

Problemas não linearmente separáveis ?



Support Vector Machines - SVM

Exemplos de Aplicações

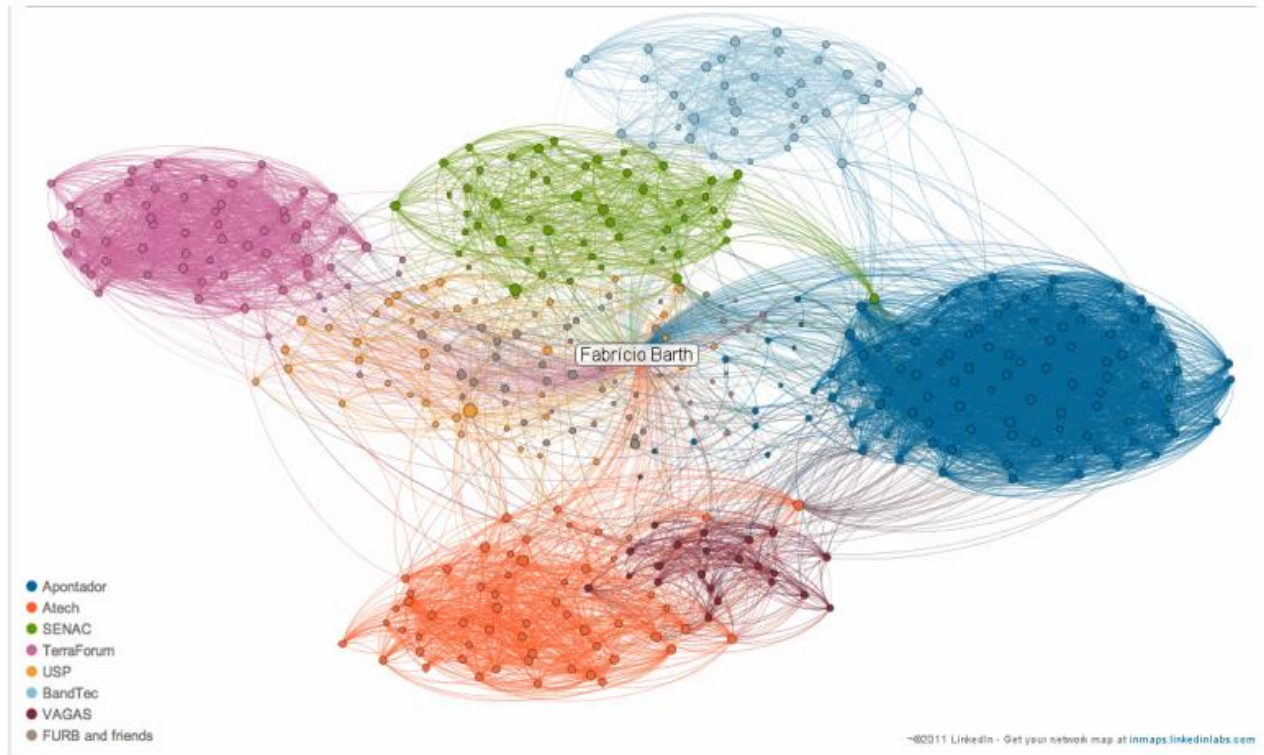
- Categorização de textos
- Análise de imagens
- Bioinformática

Exemplos de Implementações

- LIBSVM: Uma implementação eficiente do SVM, Classificação multiclases, incluindo várias interfaces com java, python, etc.
- SVM-light: Implementação mais simples, com performance inferior ao LIBSVM. Suporta somente classificação binária. Implementada em linguagem C.
- SVM-torch: implementação também em Linguagem C.
- SVM - <http://www.kernel-machines.org/>

Aprendizagem de Máquina

Exemplo: Identificação de grupos em redes sociais



Fonte: <http://fbarth.net.br/materiais/docs/aula04.pdf>

Aprendizagem de Máquina

