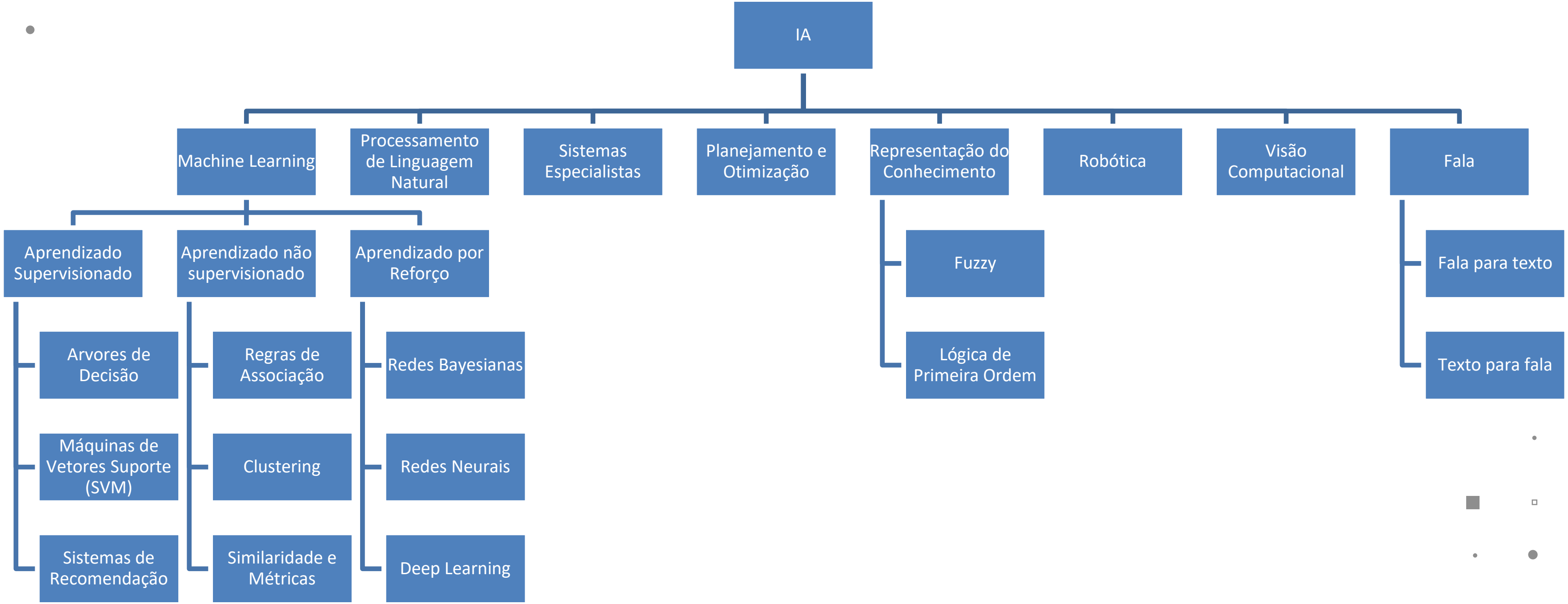
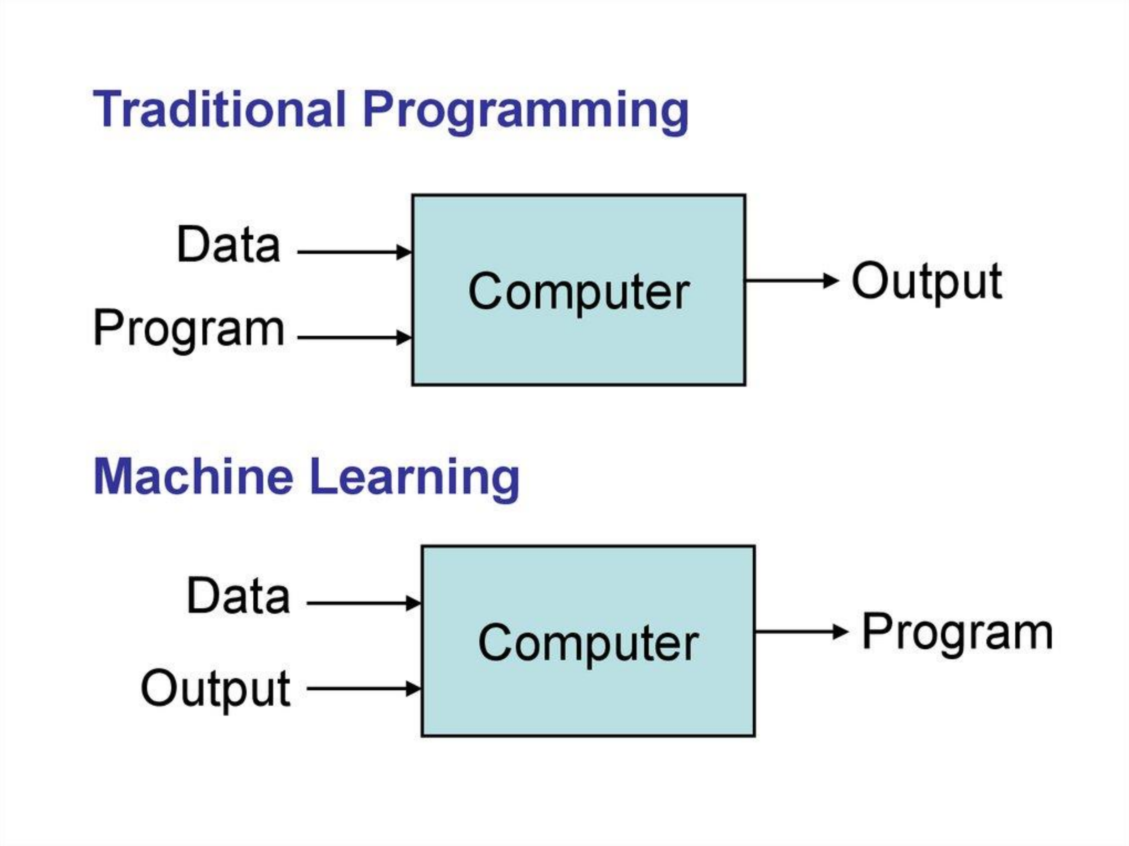
FIAP

Introdução

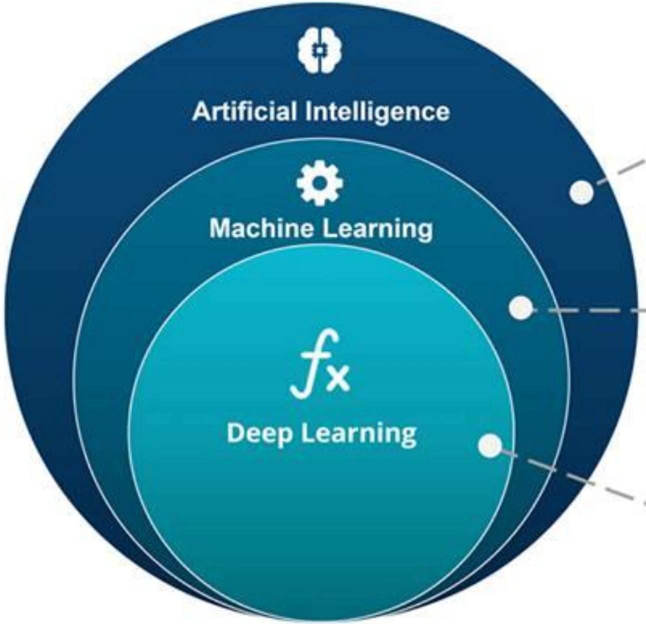
Principais Áreas da IA



Programação Tradicional vs IA



Machine e Deep Learning

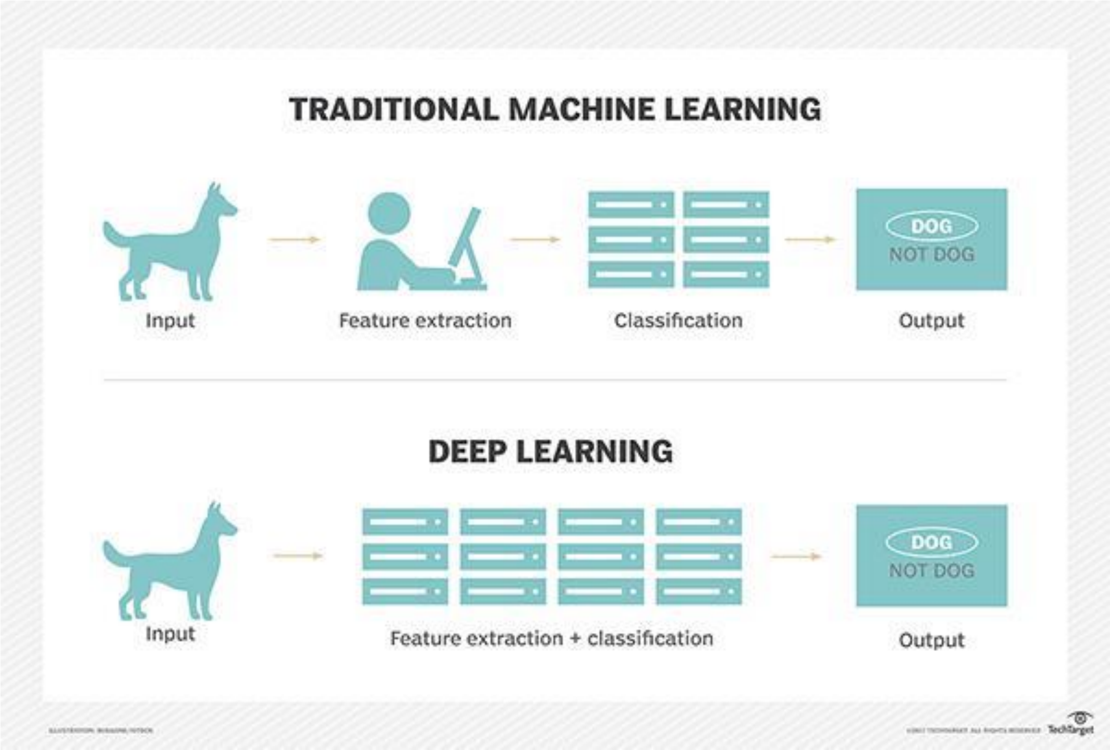


ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

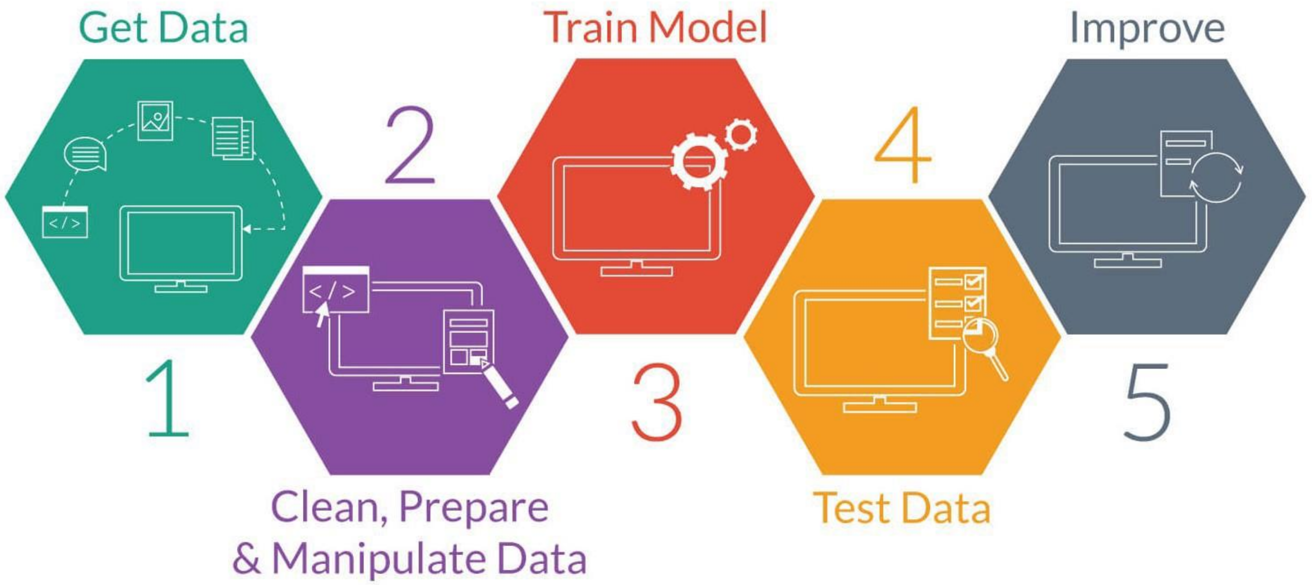
MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

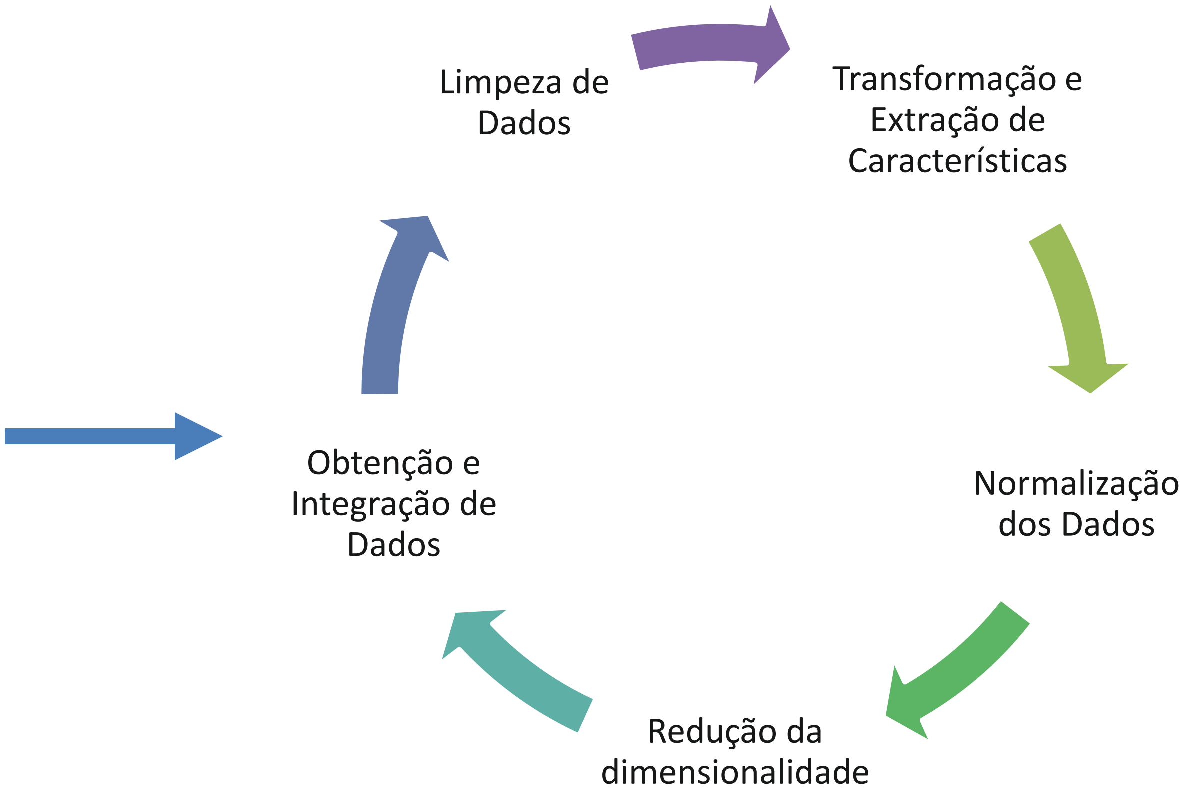
Machine e Deep Learning



Modelos de IA



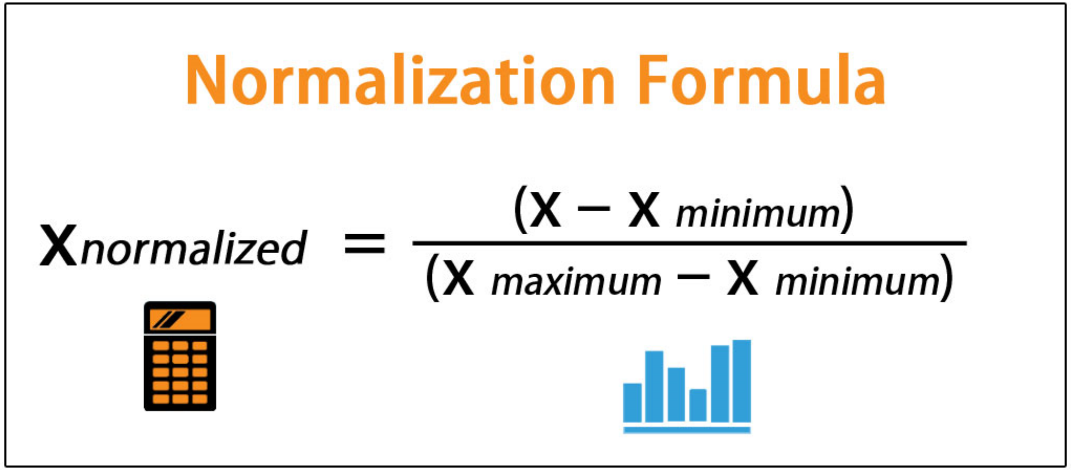
Pré-Processamento dos Dados



Pré-Processamento dos Dados

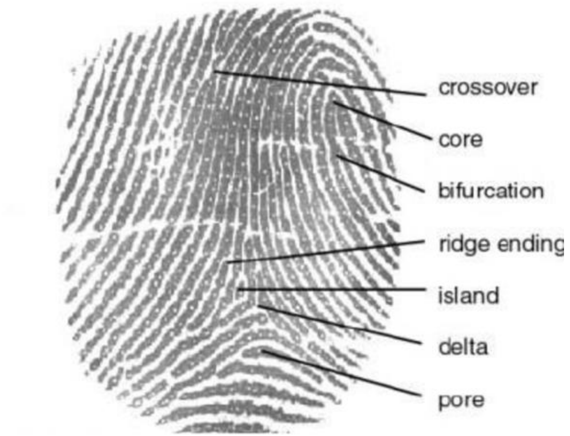
Normalização

Transformar um conjunto de dados que estão em diferentes grandezas e escalas em um conjunto de dados padronizados.



Extração de Características

Como extrair características de uma impressão digital?

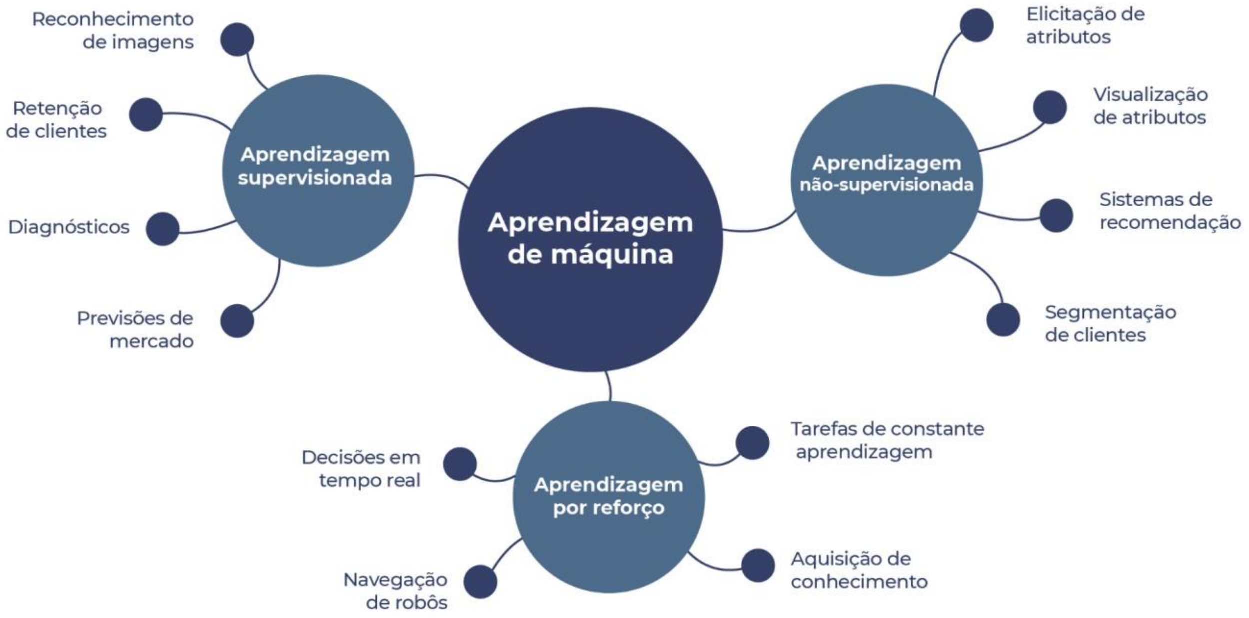


Extração de Características

Como extrair características de uma Face para reconhecimento Facial?

Demonstração

Tipo de Aprendizado e Problemas da IA

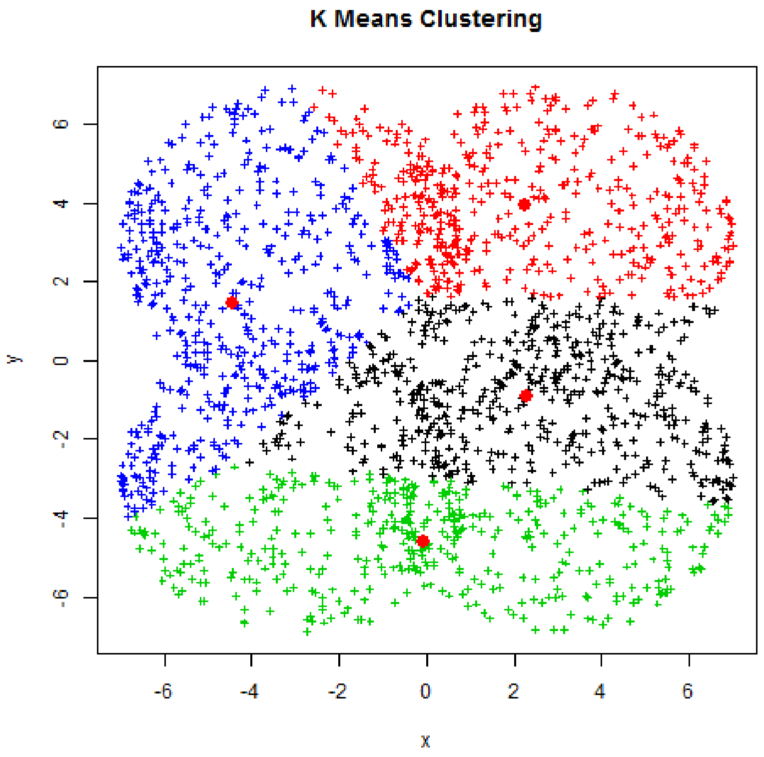


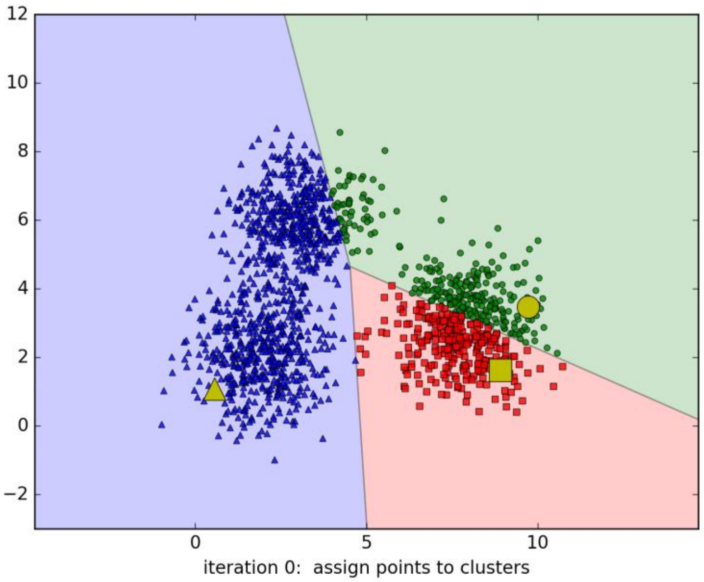
Tipos de Problema de ML e IA

Clusterização: tenta agrupar os dados mais semelhantes entre si.

Tipos de Problema de ML e IA

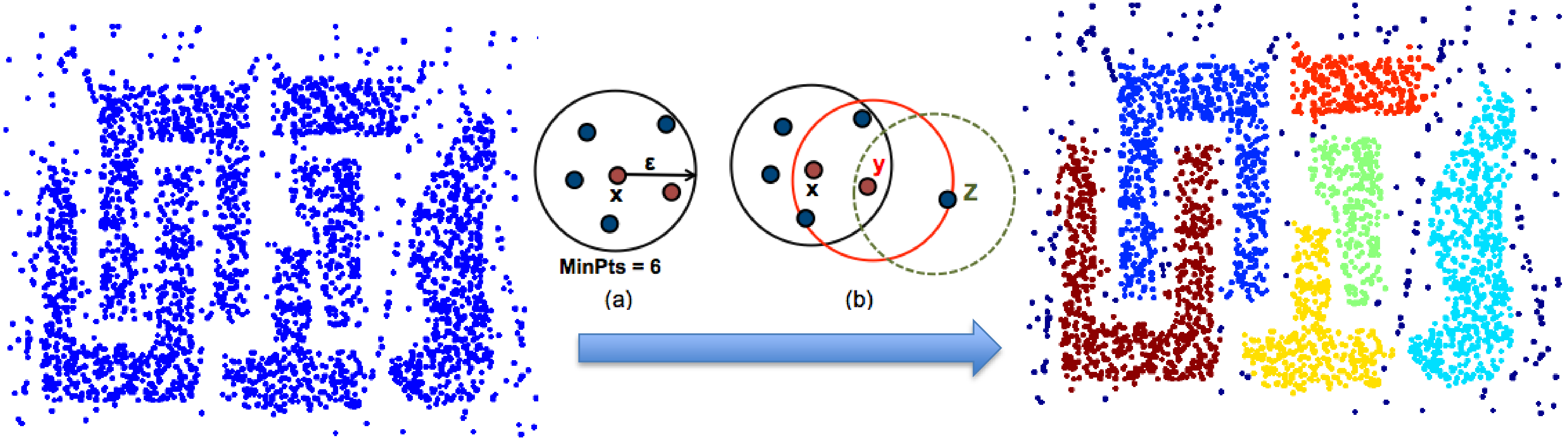
Clusterização – Algoritmo K-means:





Tipos de Problema de ML e IA

Clusterização – Algoritmo DBSCAN:

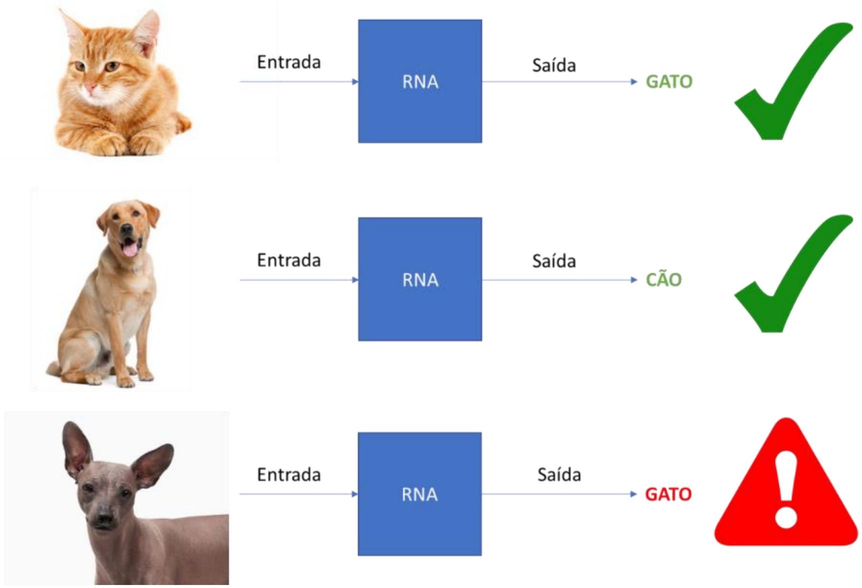


Pontos Originais

Clusters

Tipos de Problema de ML e IA

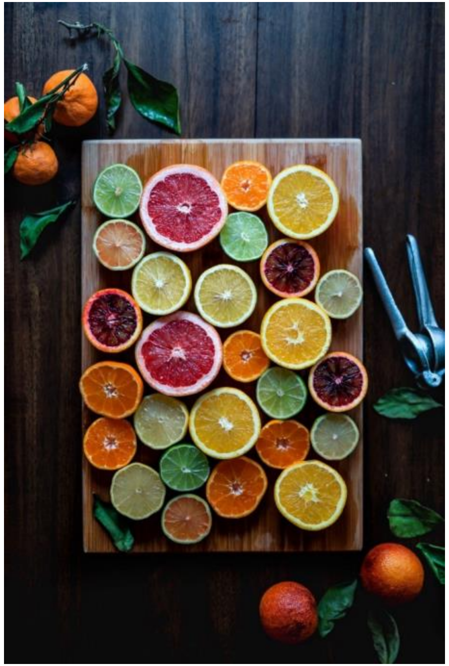
Classificação Binária: tenta predizer uma resposta simples, ex: sim ou não, cão ou gato



RNA – Rede Neural Artificial

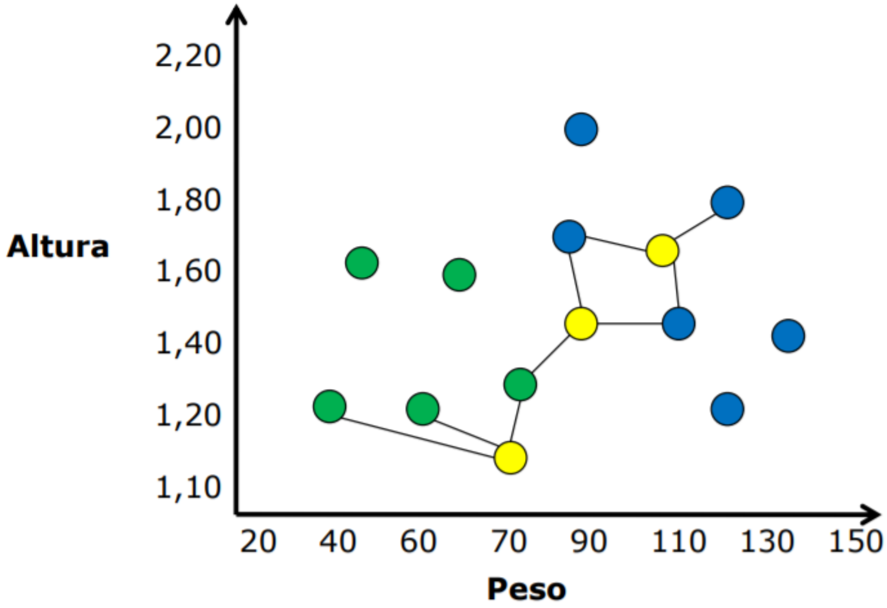
Tipos de Problema de ML e IA

Classificação Multiclasse: tenta colocar um exemplo em uma das diversas classes do problema.



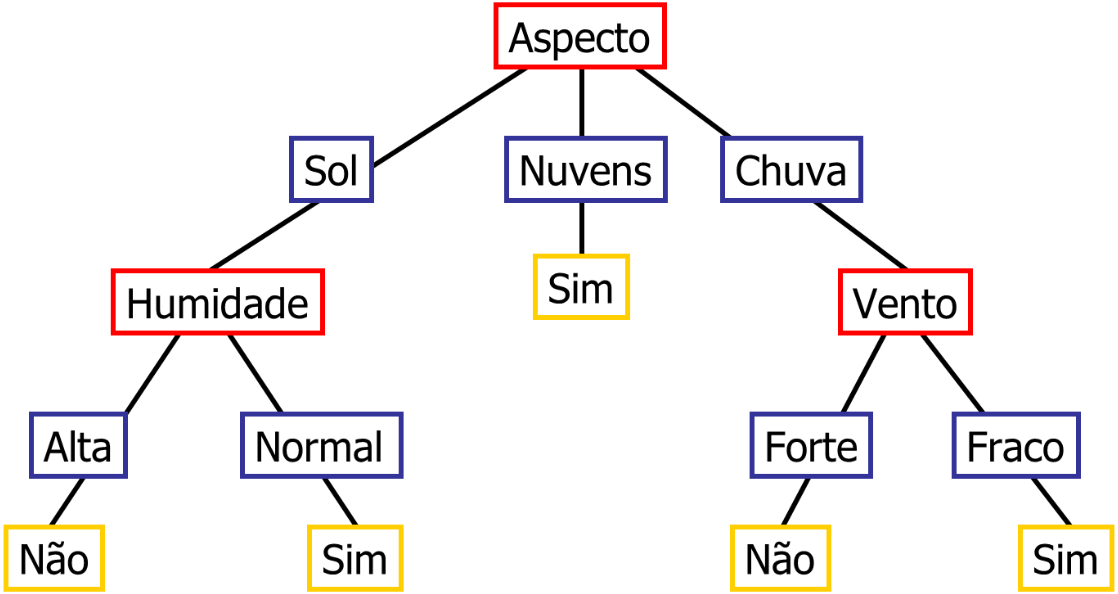
Tipos de Problema de ML e IA

Classificação – K-Nearest Neighbors



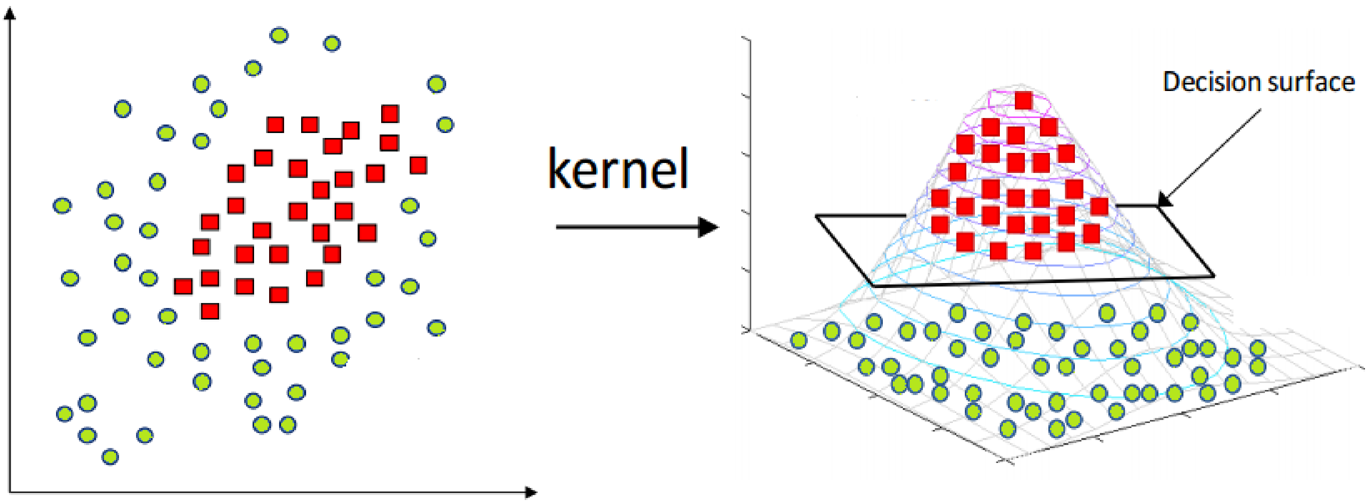
Tipos de Problema de ML e IA

Classificação - Árvores de Decisão:



Tipos de Problema de ML e IA

Classificação - Support Vector Machines:

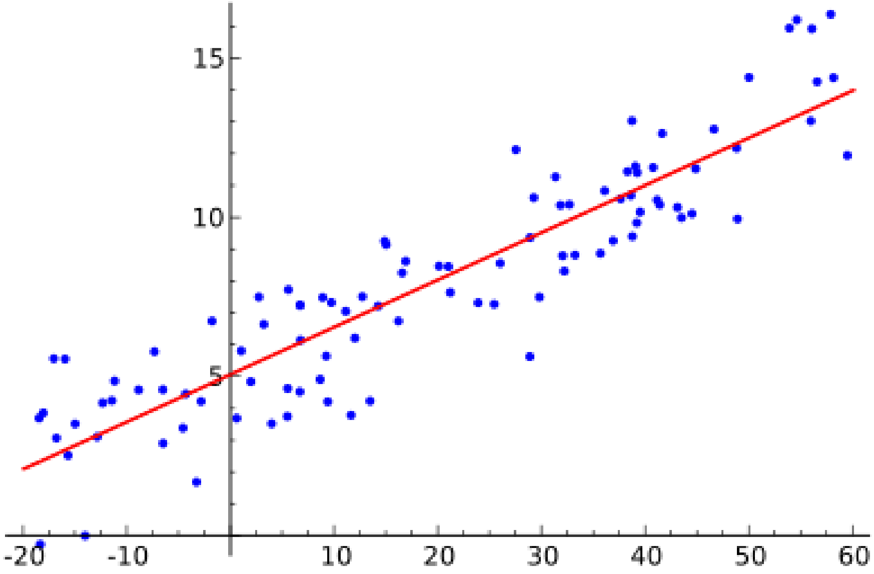


Tipos de Problema de ML e IA

Regressão: Tenta predizer um valor real.

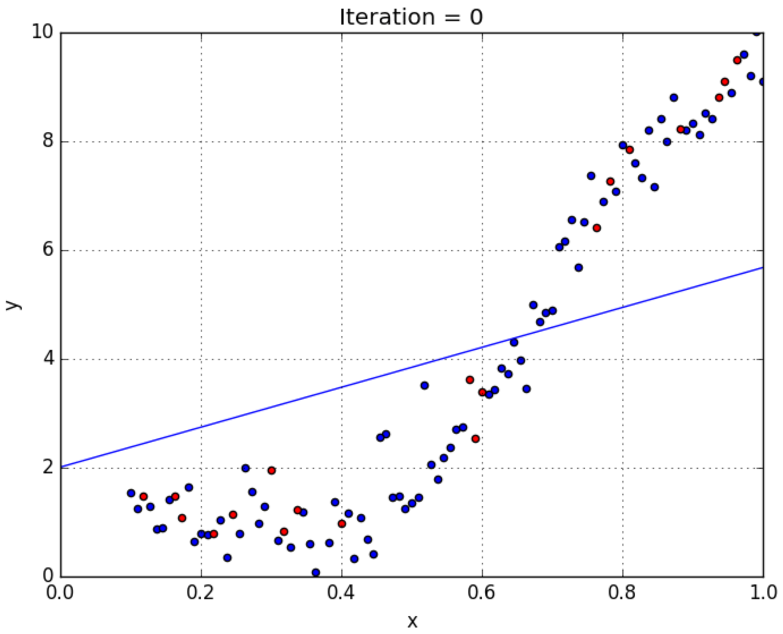
Tipos de Problema de ML e IA

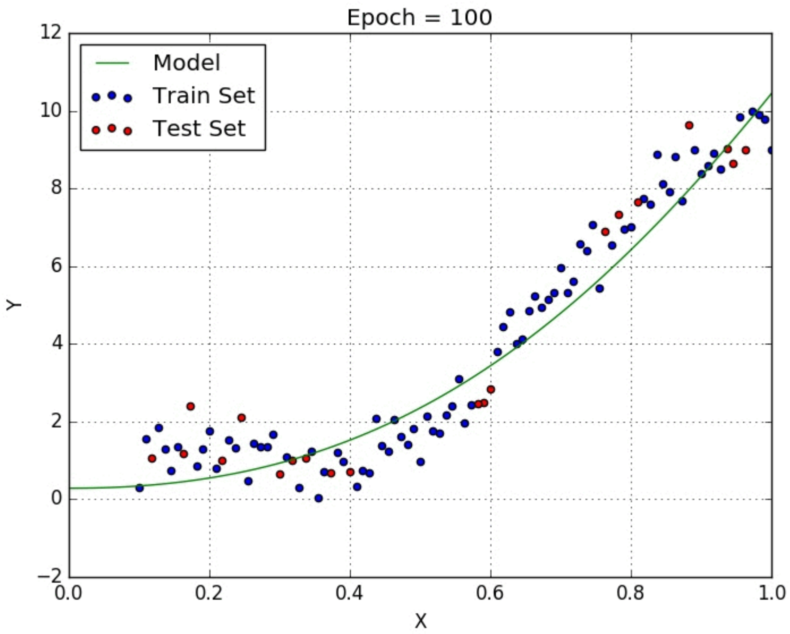
Regressão – Regressão Linear:



Tipos de Problema de ML e IA

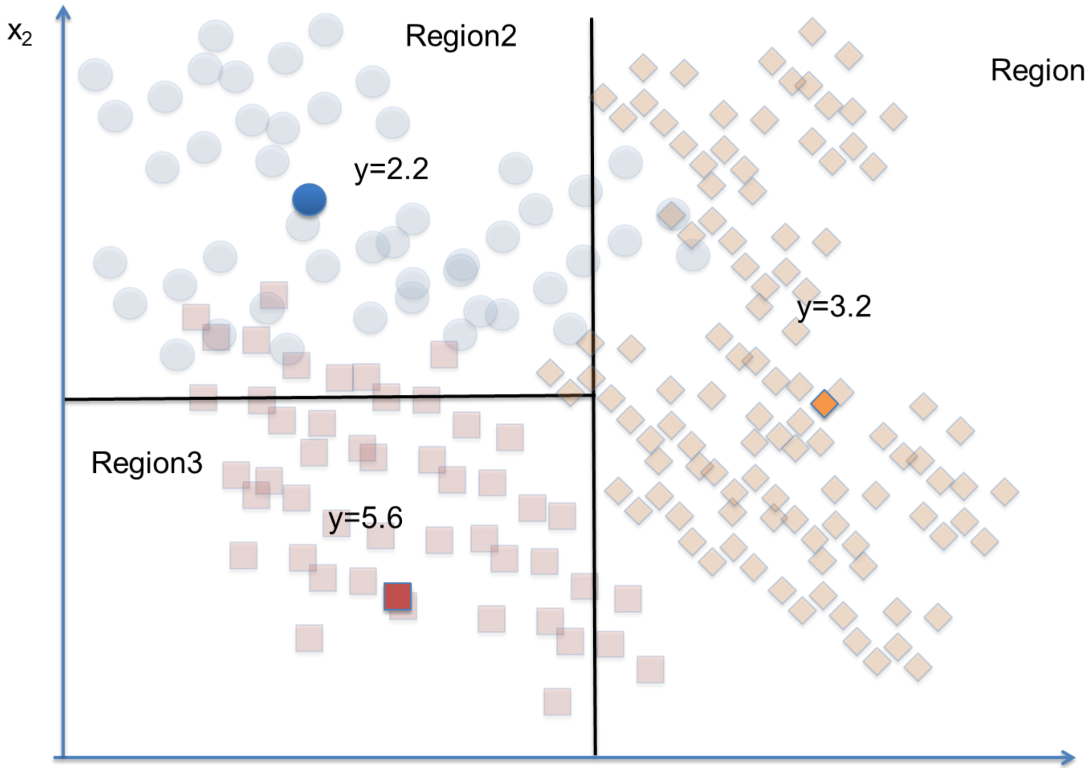
Regressão – Regressão Polinomial:





Tipos de Problema de ML e IA

Regressão – Arvores de Decisão para Regressão:

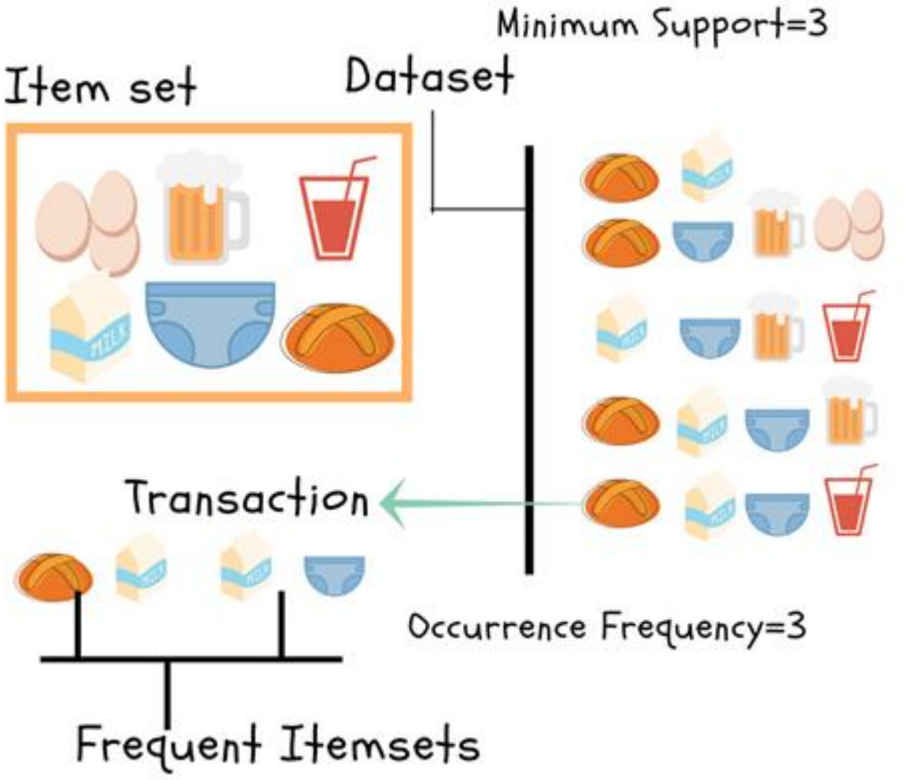


Tipos de Problema de ML e IA

Descoberta de Conhecimento

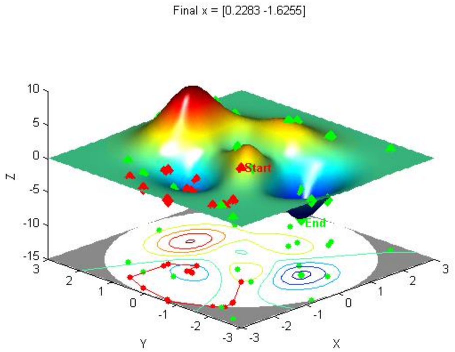
Tipos de Problema de ML e IA

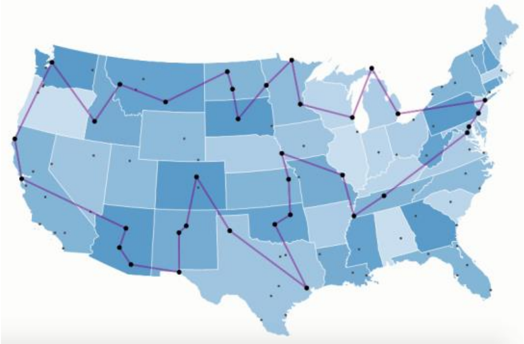
Descoberta de Conhecimento



Tipos de Problema de ML e IA

Otimização





Sistemas de Recomendação

YouTube

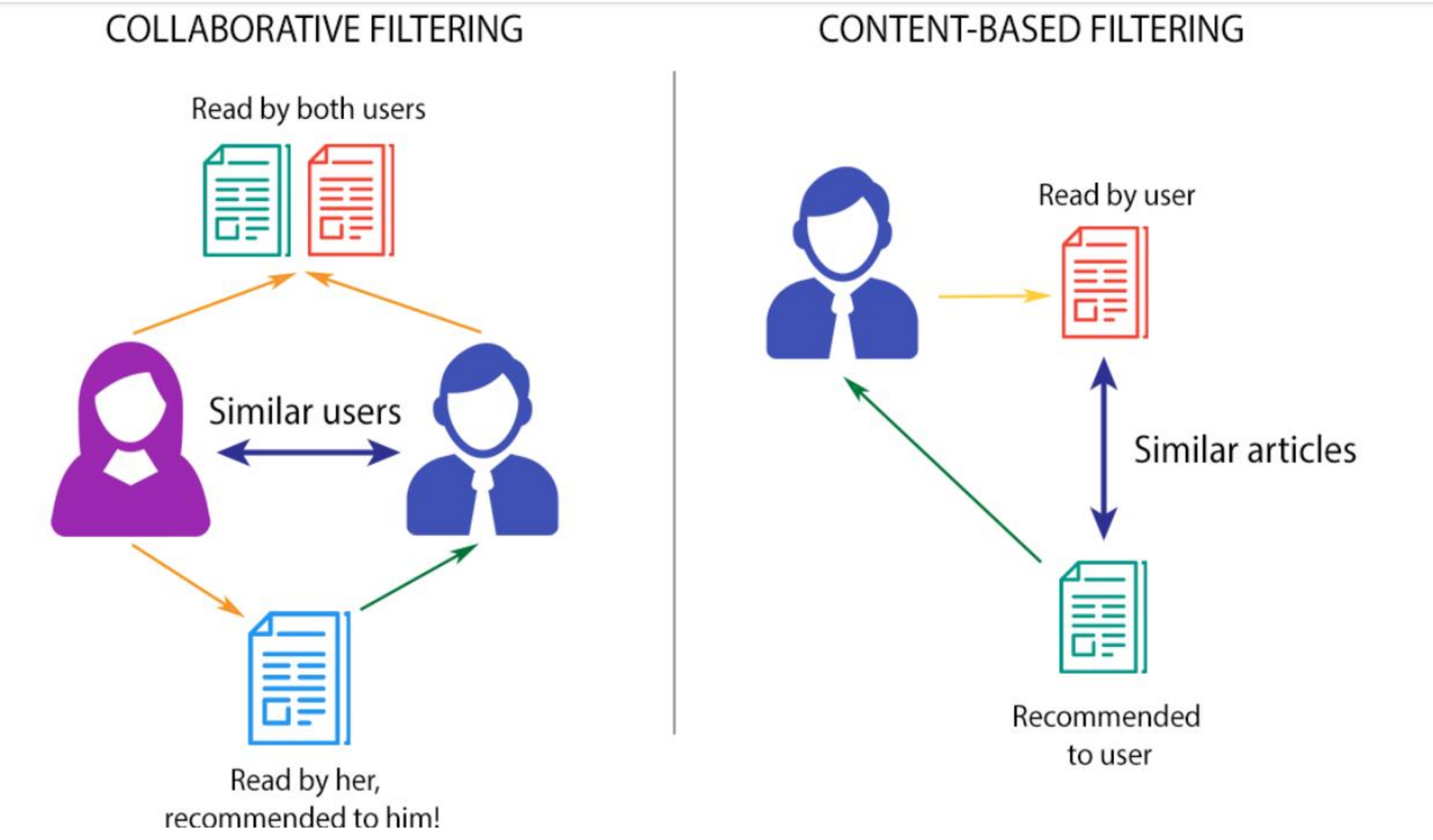
NETFLIX

Spotify

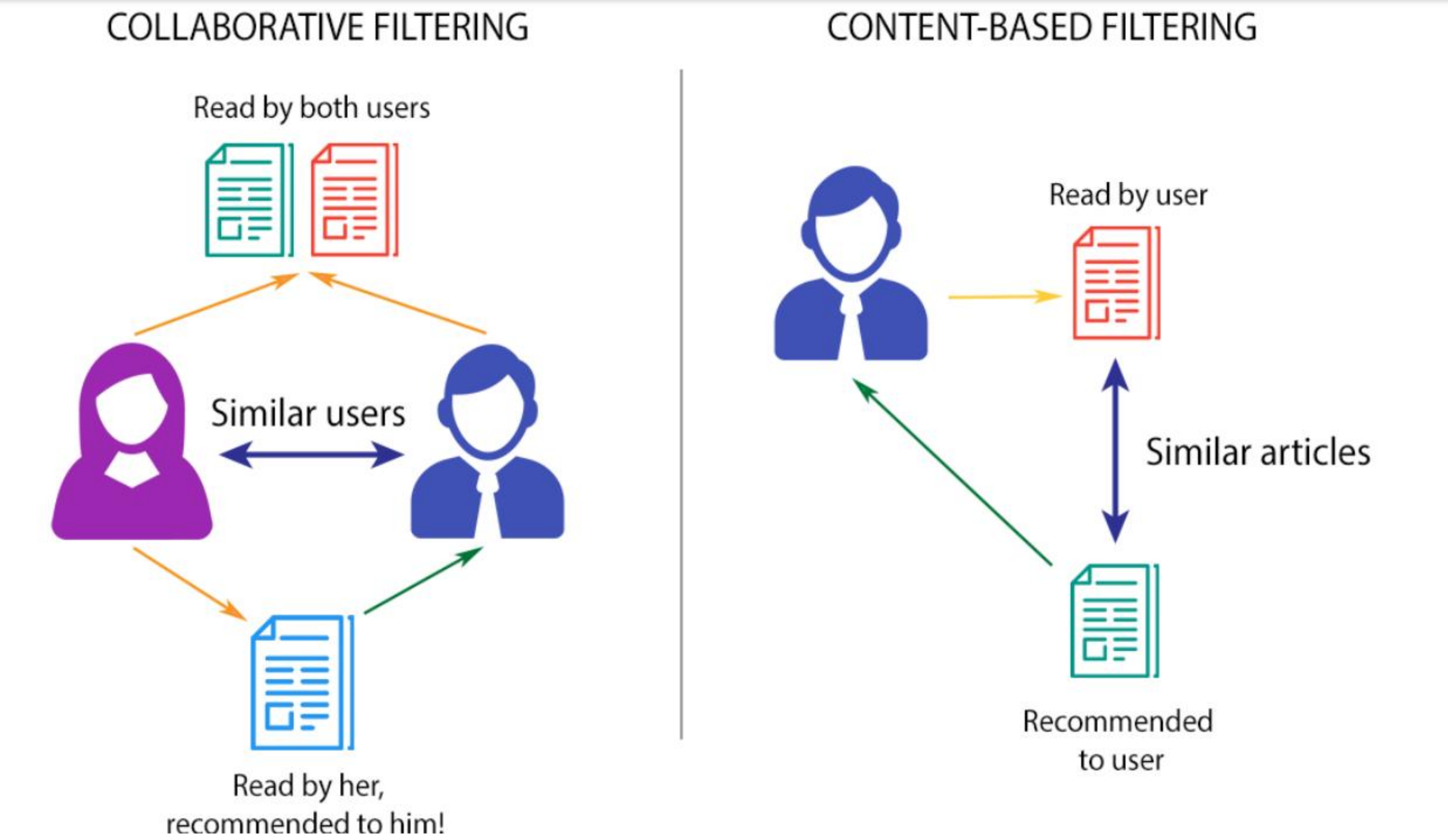
udemy

amazon.com

Sistemas de Recomendação Simples



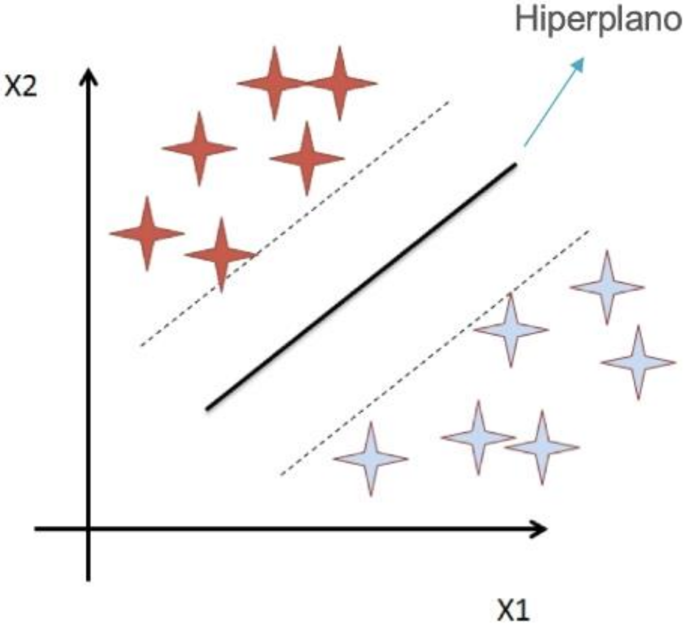
Sistemas de Recomendação Simples



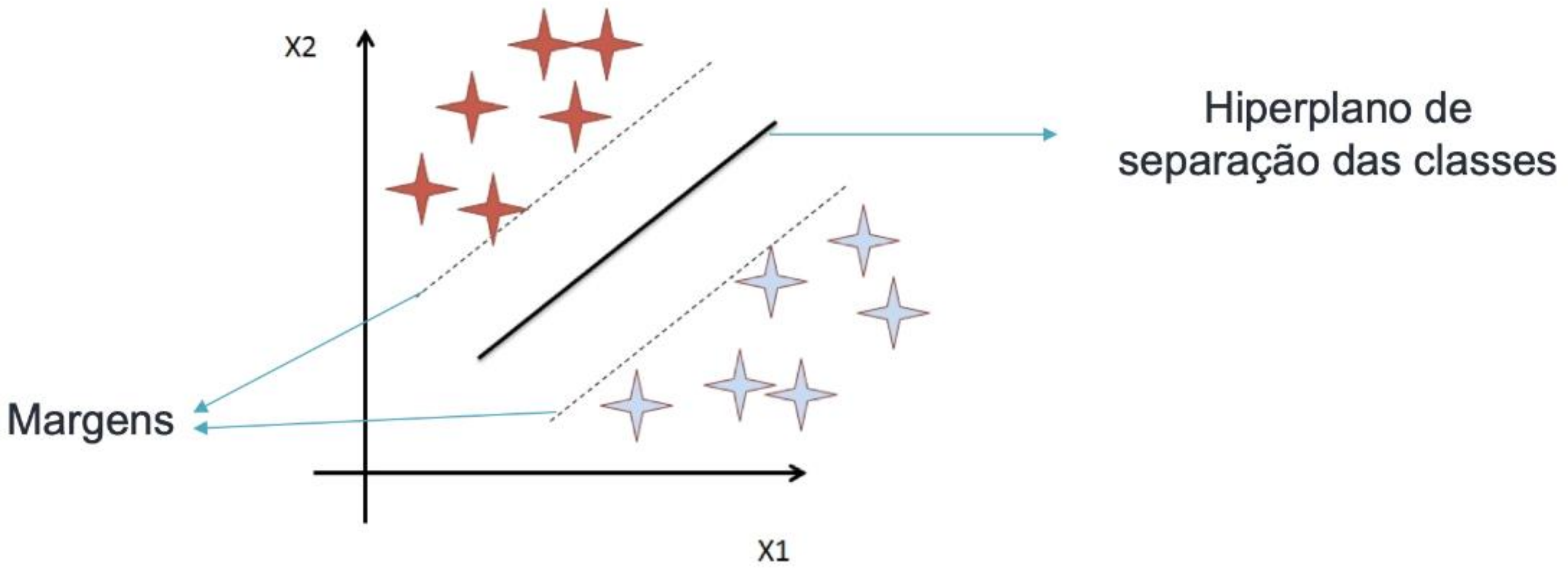
Machine Learning – SVM Suport Vector Machines

Machine Learning – SVM Suport Vector Machines

Vladimir Vapnik - 1963



Machine Learning – SVM Suport Vector Machines



Machine Learning – SVM Suport Vector Machines

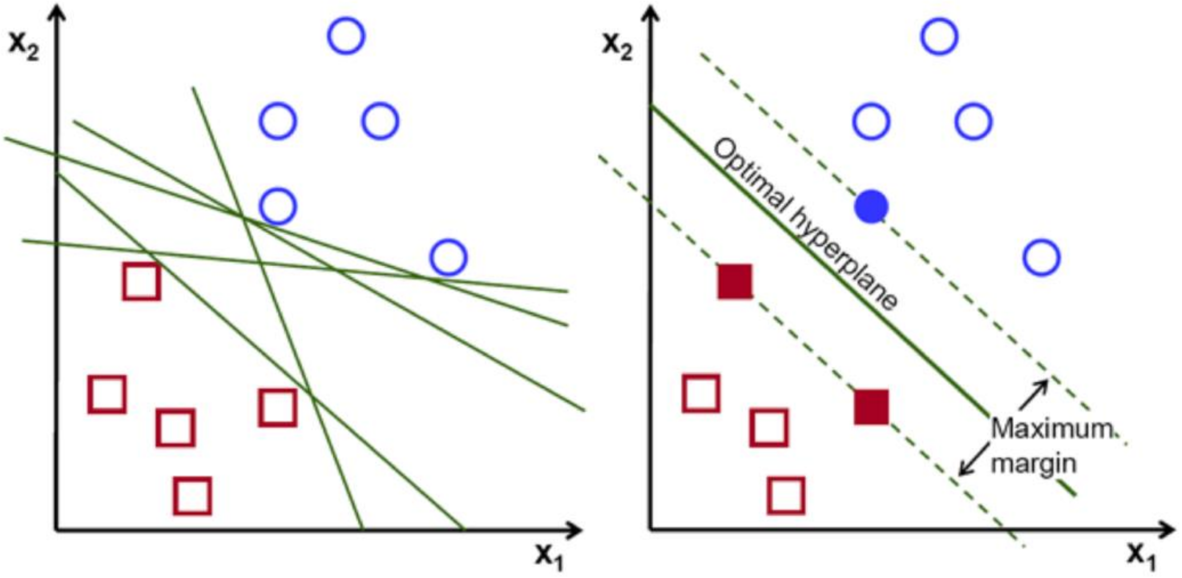
Algumas características das SVM's:

- Em caso de outliers o modelo SVM busca a melhor forma possível de classificação e, se necessário, desconsidera o outlier;
- É um classificador criado para fornecer separação linear;
- Funciona muito bem em domínios complicados, em que existe uma clara margem de separação;
- Não funciona bem em conjuntos de dados muito grandes, pois o tempo de treinamento é muito custoso;
- Não funciona bem em conjuntos de dados com grande quantidade de ruídos;
- Se as classes estiverem muito sobrepostas deve-se utilizar apenas evidências independentes.

Machine Learning – SVM Suport Vector Machines

O objetivo do algoritmo da máquina de vetores de suporte (SVM – Support Vector Machine) é encontrar um hiperplano em um espaço N-dimensional (N - o número de recursos ou atributos) que classifica distintamente os pontos de dados.

O Que São Vetores de Suporte?



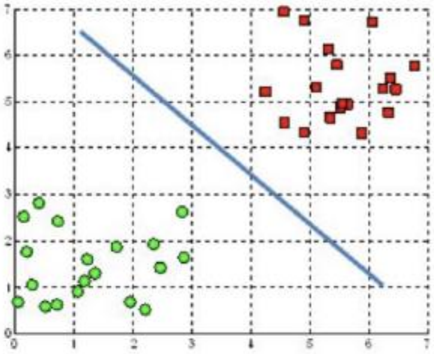
Machine Learning – SVM Suport Vector Machines

Para separar as duas classes de pontos de dados, existem muitos hiperplanos possíveis que podem ser escolhidos. Nosso objetivo é encontrar um hyperplano com a margem máxima, ou seja, a distância máxima entre os pontos de dados das duas classes. A maximização da distância da margem fornece um limite para que os pontos de dados futuros possam ser classificados com mais confiança.

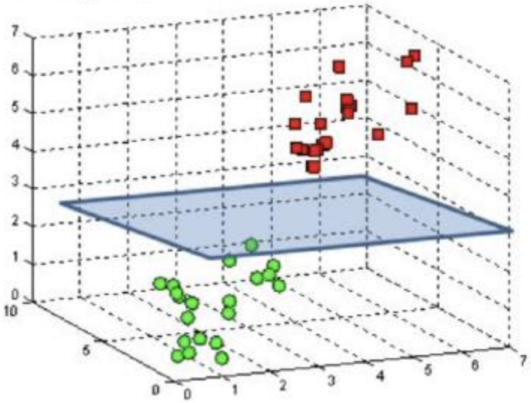
Hiperplanos são limites de decisão que ajudam a classificar os pontos de dados. A dimensão do hiperplano depende do número de recursos. Se o número de recursos de entrada for 2, o hiperplano será apenas uma linha. Se o número de recursos de entrada for 3, o hiperplano se tornará um plano bidimensional. Torna-se difícil imaginar quando o número de recursos excede 3.

Machine Learning – SVM Suport Vector Machines

A hyperplane in $\mathbb{R}^2$ is a line

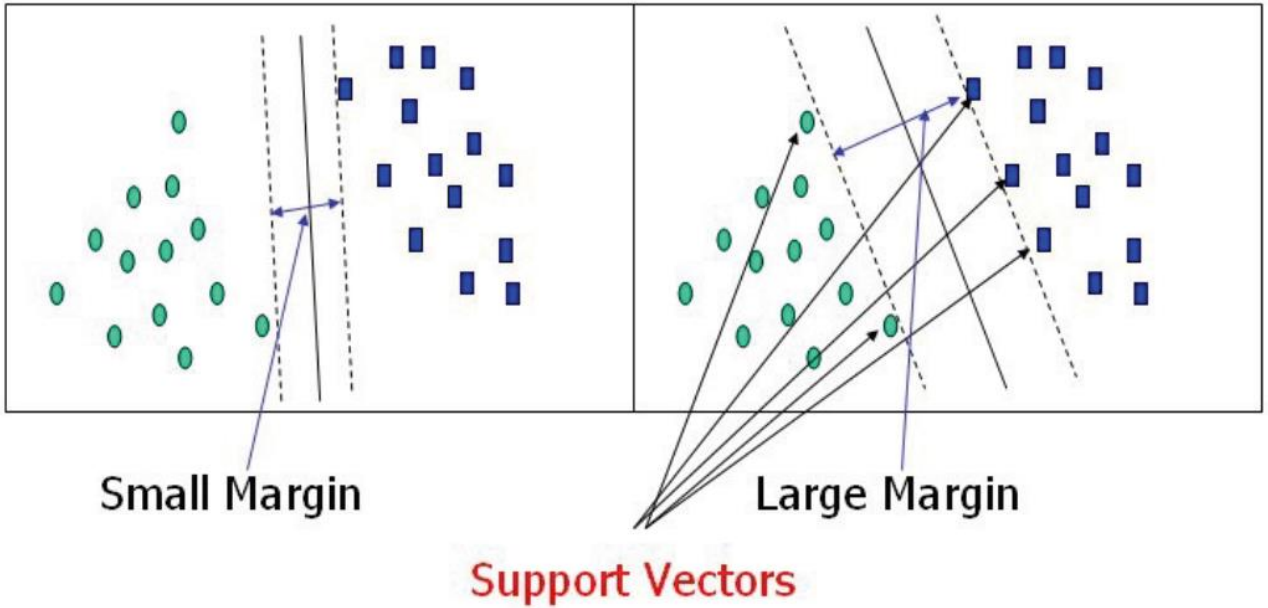


A hyperplane in $\mathbb{R}^3$ is a plane



Machine Learning – SVM Suport Vector Machines

Os vetores de suporte são pontos de dados que estão mais próximos do hiperplano e influenciam a posição e a orientação do hiperplano. Usando esses vetores de suporte, maximizamos a margem do classificador. A exclusão dos vetores de suporte alterará a posição do hiperplano. Esses são os pontos que nos ajudam a criar nosso modelo SVM.



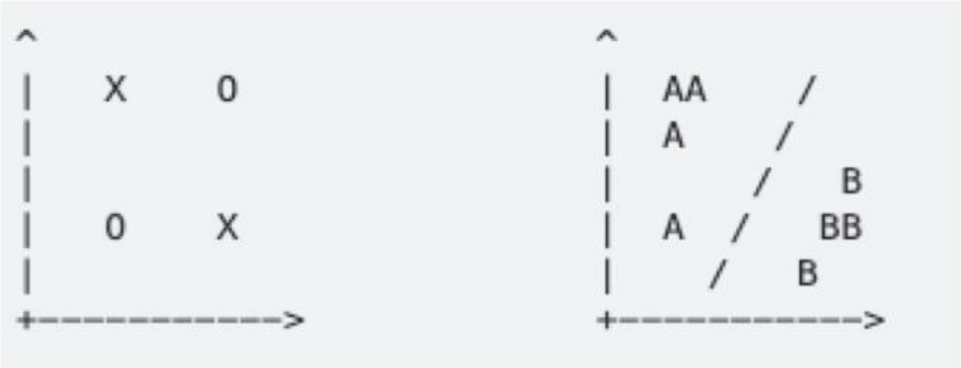
Machine Learning – SVM Suport Vector Machines

Support Vector Machines (SVM's) são modelos de aprendizagem supervisionada, que possuem algoritmos de aprendizagem que analisam dados e reconhecem padrões, utilizados para classificação e análise de regressão.

Machine Learning – SVM Suport Vector Machines

Funcionamento do Modelo SVM para dados linearmente separáveis.

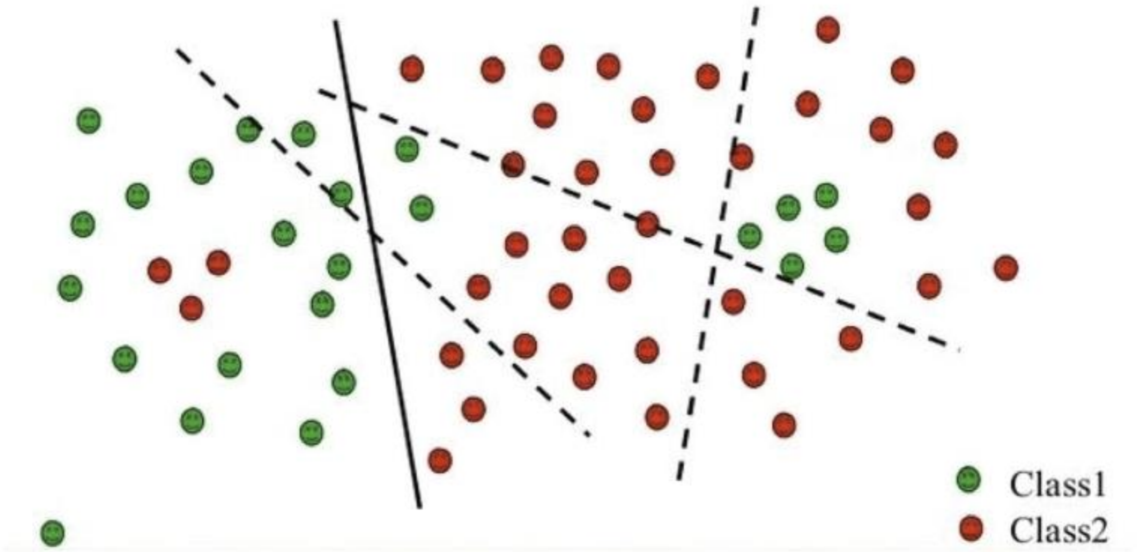
Machine Learning – SVM Suport Vector Machines



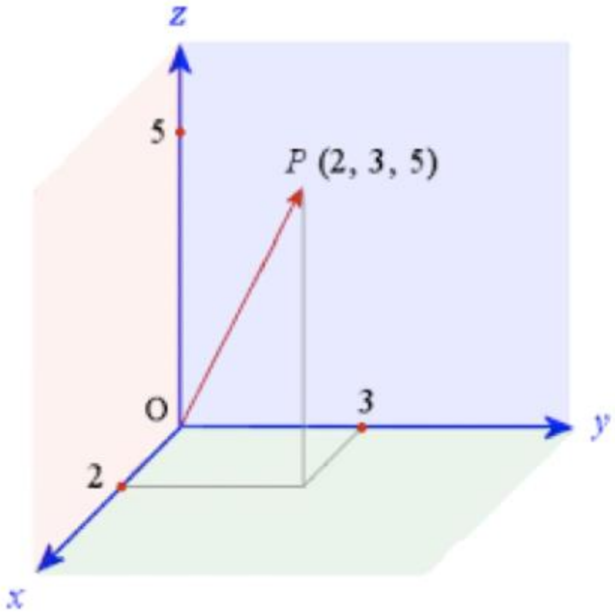
Dados Não Linearmente Separáveis

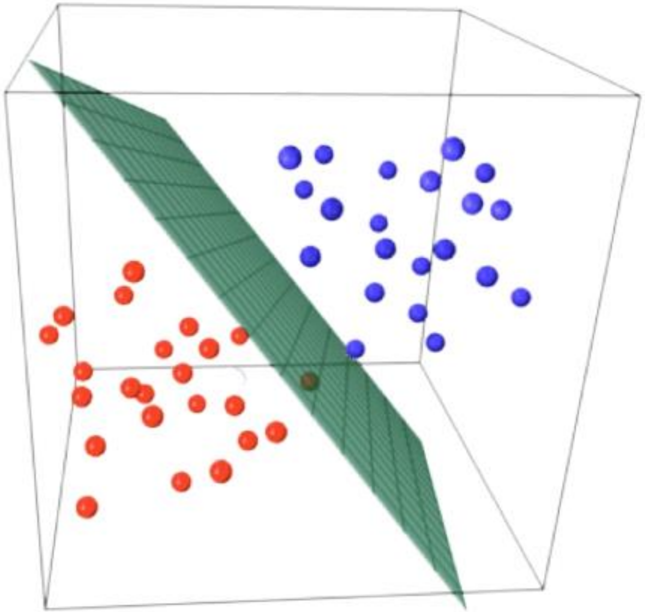
Dados Linearmente Separáveis

Machine Learning – SVM Suport Vector Machines

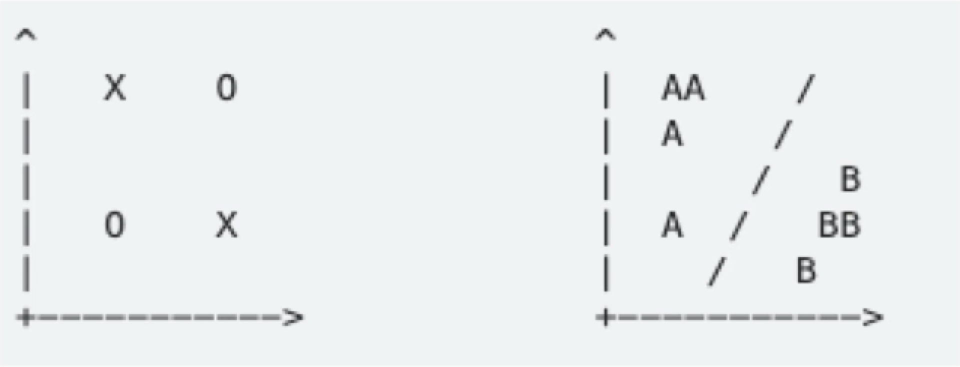


Machine Learning – SVM Suport Vector Machines





Machine Learning – SVM Suport Vector Machines



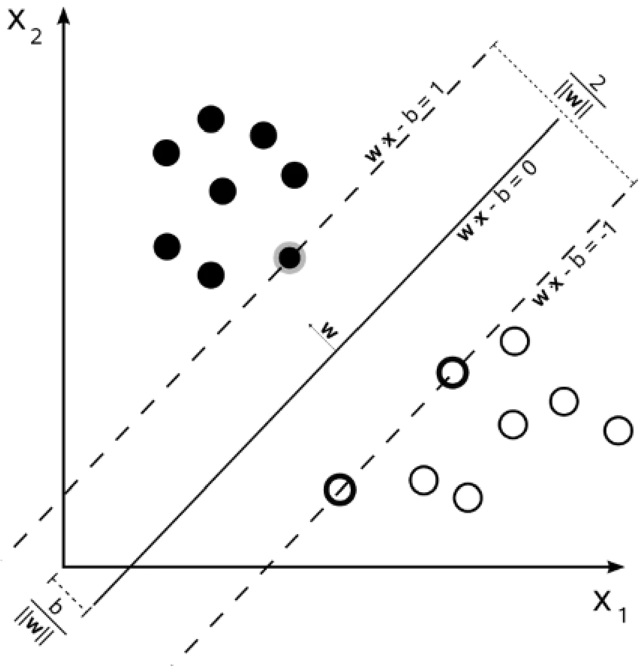
Dados Não Linearmente Separáveis

Dados Linearmente Separáveis

Precisamos de Função de Kernel Para a Separação

Machine Learning – SVM Suport Vector Machines

Uma Dose de Matemática



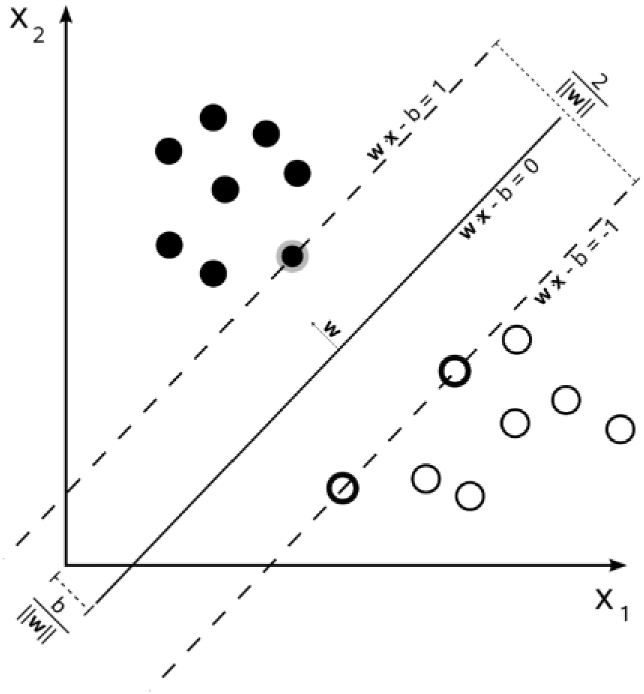
Encontrar o valor de y:

$$y^{(i)} = \begin{cases} -1 & \text{if } \mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + b \leq -1 \\ 1 & \text{if } \mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + b \geq 1 \end{cases}$$

Distância Mínima Entre os VS:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \frac{\|\mathbf{w}\|}{2} + C \sum_{i=1}^{N} \xi^{(i)}, & \\ s.t. \quad & y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + b) \geq 1 - \xi^{(i)}, & \forall i \in \{1, \ldots, N\} \\ & \xi^{(i)} \geq 0, & \forall i \in \{1, \ldots, N\} \end{aligned}$$

Machine Learning – SVM Suport Vector Machines



Distância Mínima Entre os VS:

$$\min_{\mathbf{w},b} \quad \frac{\|\mathbf{w}\|}{2} + C\sum_{i=1}^{N} \xi^{(i)},$$

$$s.t. \quad y^{(i)}(\mathbf{w}^{\mathbf{T}}\mathbf{x}^{(i)} + b) \geq 1 - \xi^{(i)}, \qquad \forall i \in \{1, \dots, N\}$$

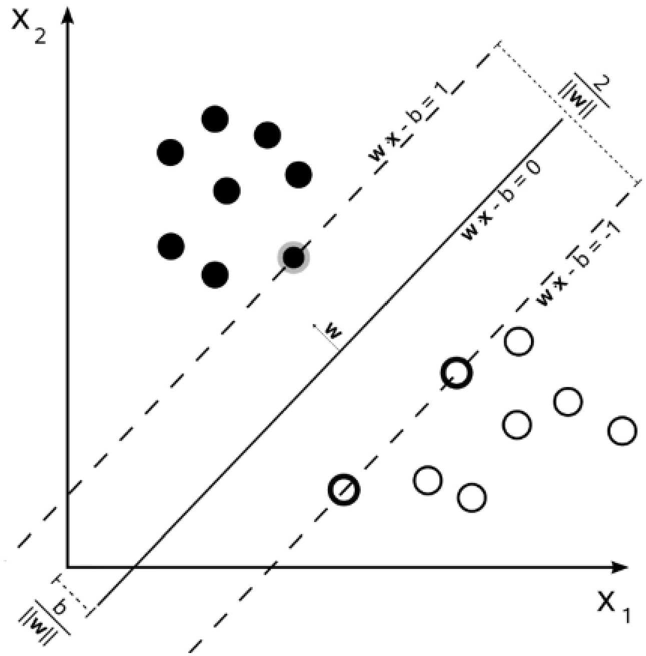
$$\xi^{(i)} \geq 0, \qquad \forall i \in \{1, \dots, N\}$$

Maximizar a Distância Mínima (Otimização):

$$\max_{\alpha} \quad \sum_{i=1}^{N} \alpha^{(i)} - \sum_{i=1}^{N}\sum_{j=1}^{N} \left(y^{(i)}\alpha^{(i)}\phi\left(\mathbf{x}^{(i)}\right)^{T}\phi\left(\mathbf{x}^{(j)}\right) y^{(j)}\alpha^{(j)} \right)$$

$$s.t. \quad 0 \leq \alpha^{(i)} \leq C,$$

Machine Learning – SVM Suport Vector Machines



Maximizar a Distância Mínima (Otimização):

$$\max_{\alpha} \quad \sum_{i=1}^{N} \alpha^{(i)} - \sum_{i=1}^{N} \sum_{j=1}^{N} \left(y^{(i)} \alpha^{(i)} \underbrace{\phi\left(\mathbf{x}^{(i)}\right)^{T} \phi\left(\mathbf{x}^{(j)}\right)}_{} y^{(j)} \alpha^{(j)} \right)$$

$$s.t. \quad 0 \leq \alpha^{(i)} \leq C,$$

Inner Product
O Kernel Trick é este mapeamento.

Coeficiente aprendido durante o treinamento, um para i e outro para j.

Machine Learning – SVM Suport Vector Machines

O produto escalar entre dois vetores (dot product) mostra como os vetores são "semelhantes". Se os vetores representam pontos no seu conjunto de dados, o produto escalar informa se eles são semelhantes ou não.

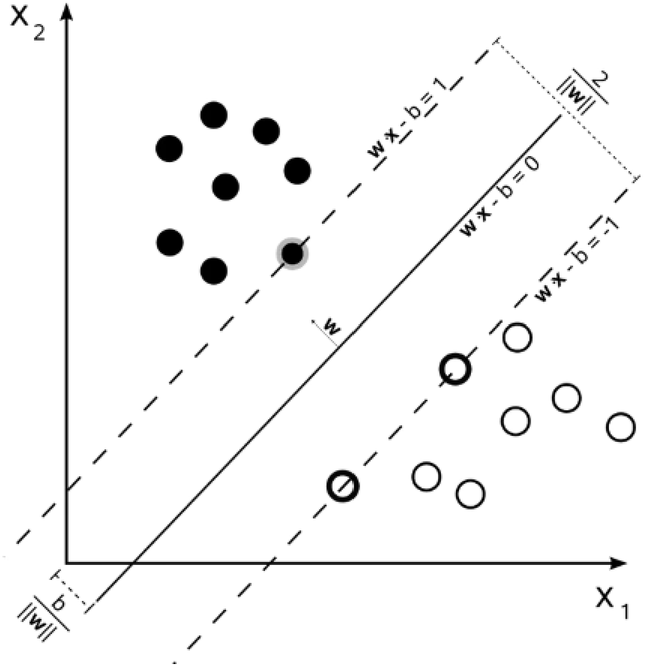
Mas, em alguns (muitos) casos, o produto escalar não é a melhor métrica de similaridade.

Por exemplo:

Talvez os pontos com produto escalar baixo sejam semelhantes por outras razões. Você pode ter itens de dados que não estão bem representados como pontos ou pode não haver separação linear.

Então, em vez de usar o produto escalar, você usa um "kernel", que é apenas uma função que recebe dois pontos e fornece uma medida de sua similaridade. O SVM aplica esse conceito que é denominado Truque do Kernel (Kernel Trick).

Machine Learning – SVM Suport Vector Machines



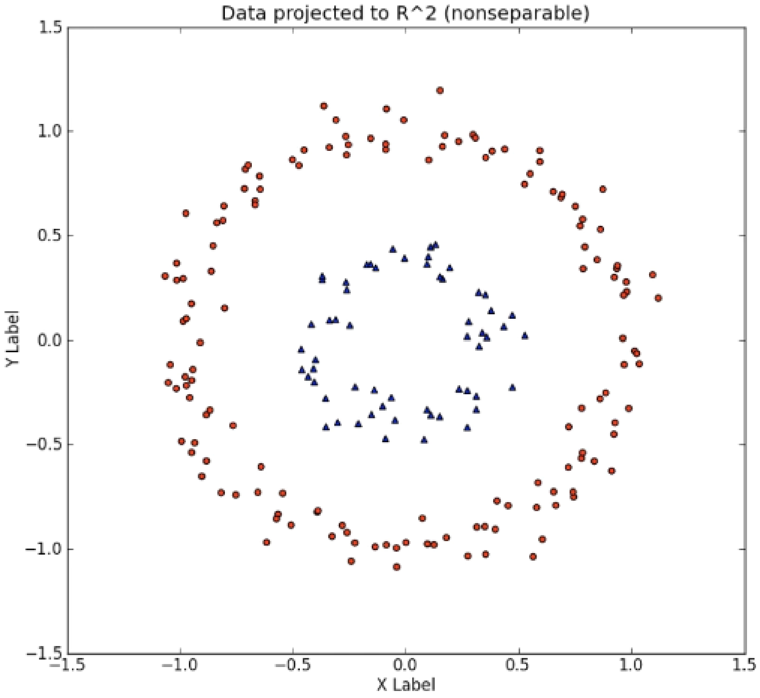
Previsões do Modelo:

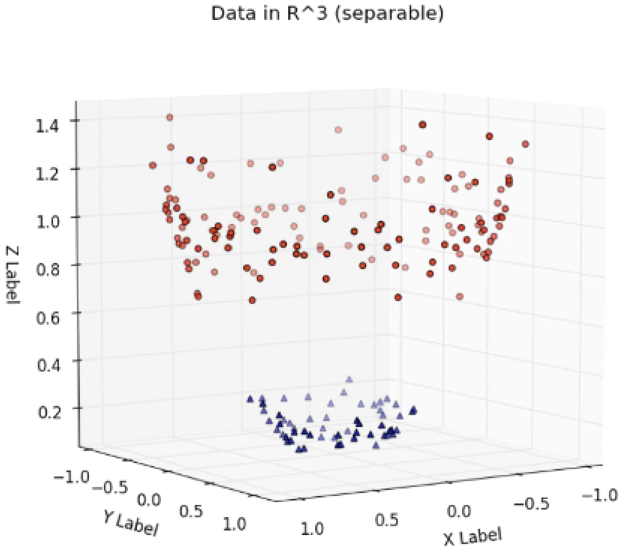
$$
\begin{aligned}
y^{\text{test}} &= \operatorname{sign}\left(\mathbf{w}^{\mathbf{T}}\phi\left(\mathbf{x}^{\text{test}}\right)+b\right) \\
&= \operatorname{sign}\left(\sum_{i=1}^{N}\alpha^{(i)}y^{(i)}\phi\left(x^{(i)}\right)^{T}\phi\left(\mathbf{x}^{\text{test}}\right)+b\right)
\end{aligned}
$$

Machine Learning – SVM Suport Vector Machines

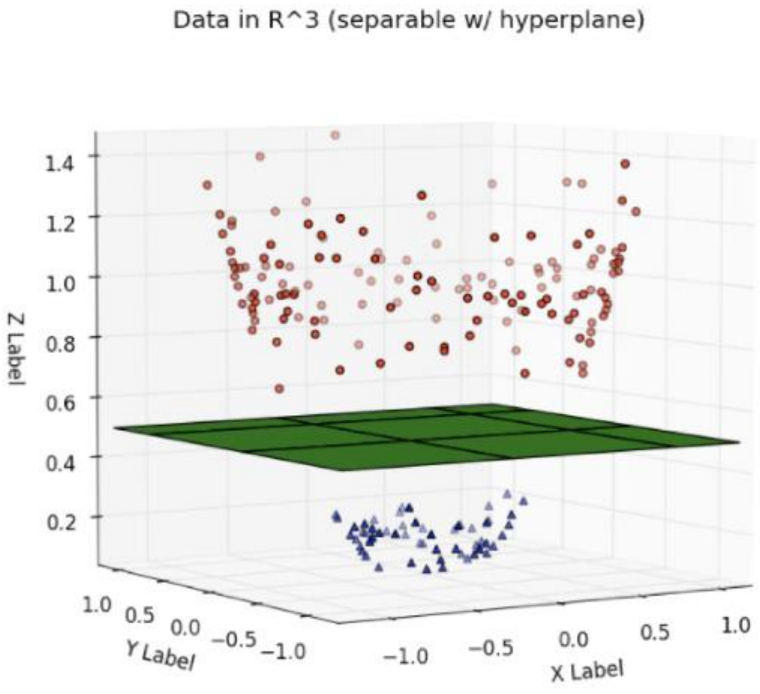
Funcionamento do Modelo SVM para dados NÃO linearmente separáveis.

Machine Learning – SVM Suport Vector Machines

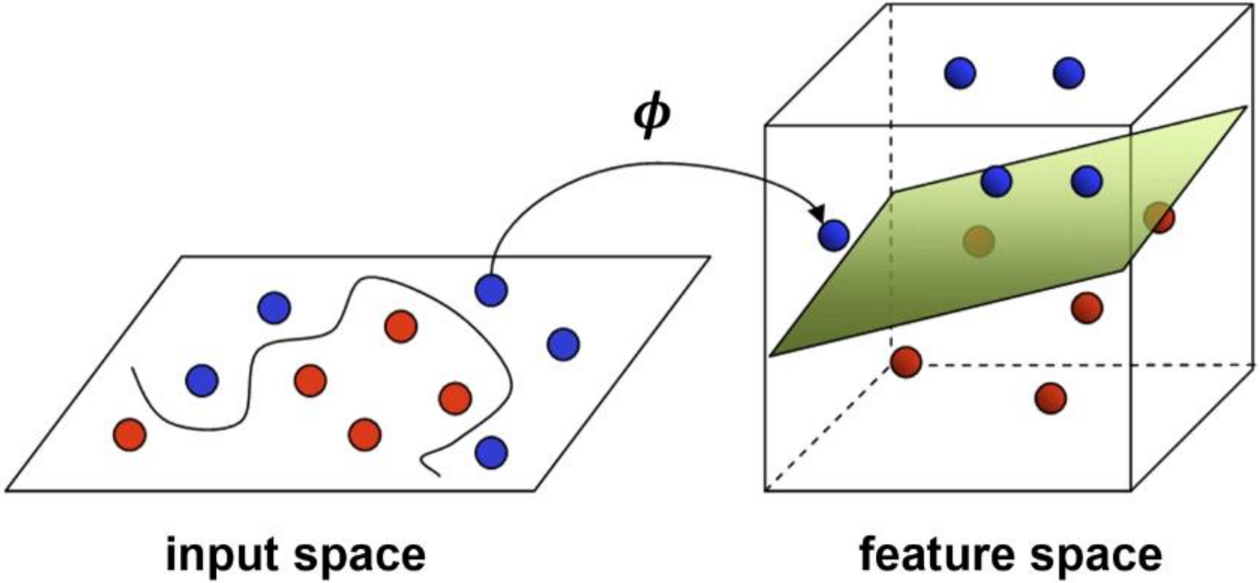




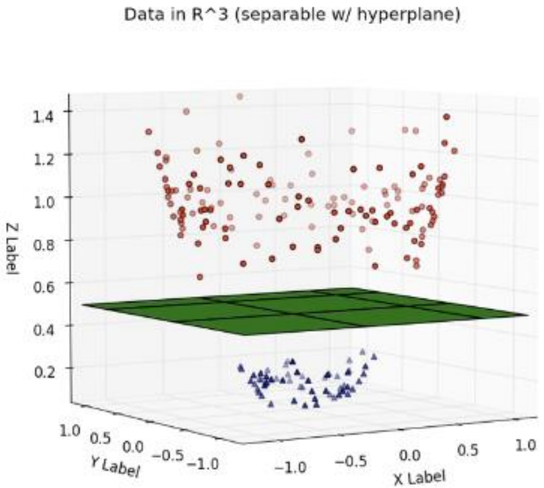
Machine Learning – SVM Suport Vector Machines

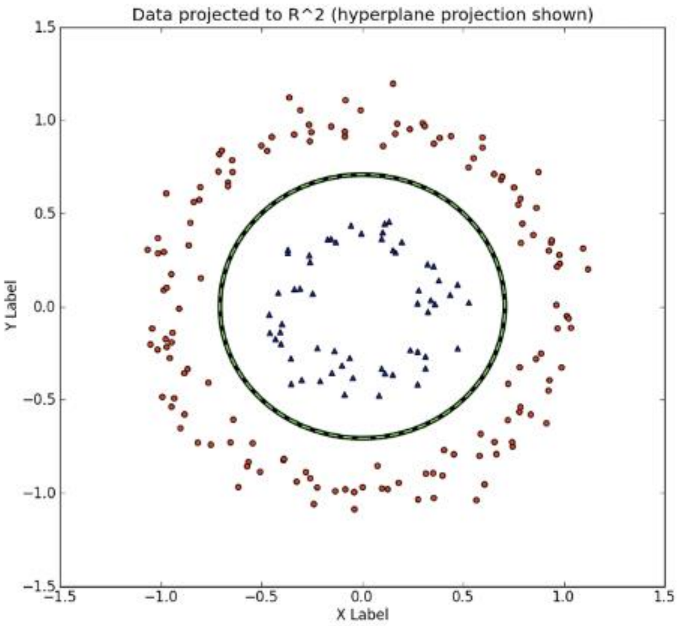


Machine Learning – SVM Suport Vector Machines



Machine Learning – SVM Suport Vector Machines





Machine Learning – SVM Suport Vector Machines

SVM's com Margens Rígidas

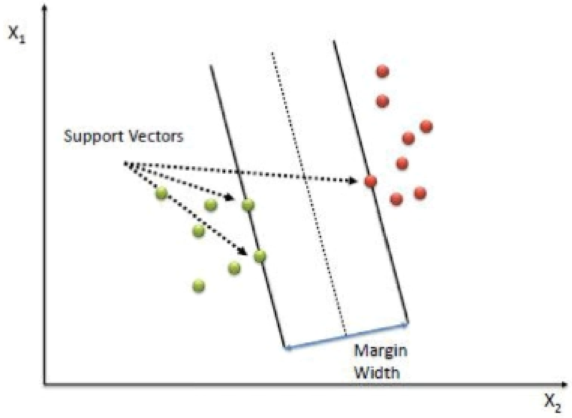
x

SVM's com Margens Flexíveis

Machine Learning – SVM Suport Vector Machines

As máquinas de vetores de suporte (chamadas SVMs) são um algoritmo de aprendizado supervisionado que pode ser usado para problemas de classificação e regressão como classificação de vetores de suporte (SVC) e regressão de vetores de suporte (SVR).

Machine Learning – SVM Suport Vector Machines

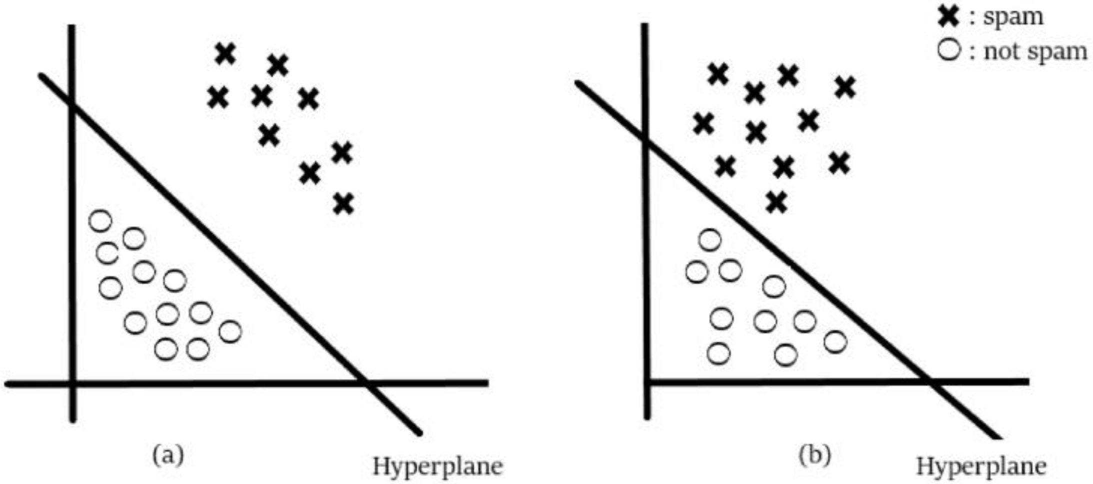


Os pontos mais próximos ao hiperplano são chamados de pontos do vetor de suporte e a distância dos vetores do hiperplano é chamada de margem.

A intuição básica a ser desenvolvida aqui é que quanto mais pontos SV adicionais, do hiperplano, maior a probabilidade de classificar corretamente os pontos em suas respectivas regiões ou classes. Os pontos SV são muito críticos na determinação do hiperplano porque se a posição dos vetores muda, a posição do hiperplano é alterada.

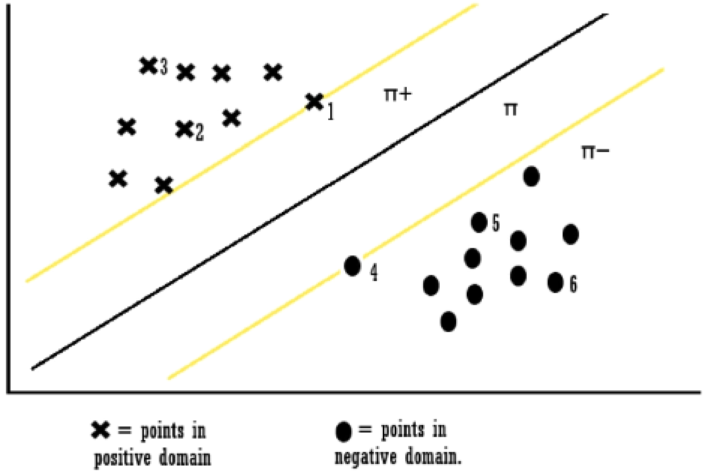
Tecnicamente, esse hiperplano também pode ser chamado de hiperplano de maximização de margem.

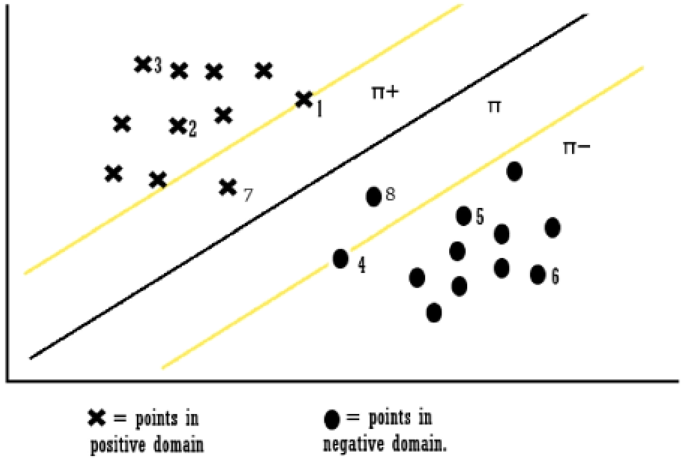
Machine Learning – SVM Suport Vector Machines



Machine Learning – SVM Suport Vector Machines

Margens Rígidas





Machine Learning – SVM Suport Vector Machines

Margens Rígidas

Se os pontos são linearmente separáveis, apenas o nosso hiperplano é capaz de distinguir entre eles e se algum erro for introduzido (outliers por exemplo), não será possível separá-los.

Esse tipo de SVM é chamado SVM de Margem Rígida (já que temos restrições muito rígidas para classificar corretamente cada ponto de dados).

Machine Learning – SVM Suport Vector Machines

Margens Flexíveis

Basicamente, consideramos que os dados são linearmente separáveis e isso pode não ser o caso no cenário da vida real.

Precisamos de uma atualização para que nossa função possa pular alguns valores discrepantes e poder classificar pontos quase linearmente separáveis. Por esse motivo, apresentamos uma nova variável Slack (ξ) chamada Xi.

Distância Mínima Entre os VS:

$$
\begin{aligned}
\min_{\mathbf{w},b} \quad & \frac{\|\mathbf{w}\|}{2} + C\sum_{i=1}^{N} \xi^{(i)}, \\
s.t. \quad & y^{(i)}(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + b) \geq 1 - \xi^{(i)}, \qquad \forall i \in \{1, \ldots, N\} \\
& \xi^{(i)} \geq 0, \qquad \forall i \in \{1, \ldots, N\}
\end{aligned}
$$

Machine Learning – SVM Suport Vector Machines

Margens Flexíveis

Se ξi = 0, os pontos podem ser considerados corretamente classificados.
Senão, se ξi > 0, pontos são classificados incorretamente.

Portanto, se ξi > 0 significa que Xi (variáveis) está na dimensão incorreta, podemos pensar em ξi como um termo de erro associado a Xi (variável).

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

Machine Learning – SVM Suport Vector Machines

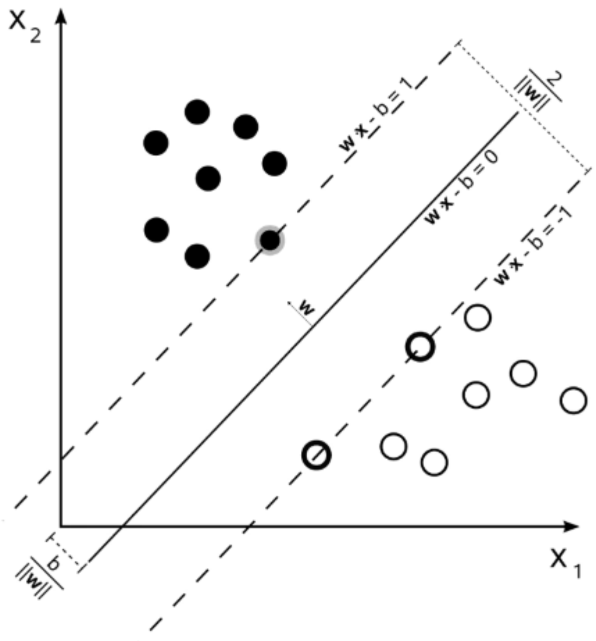
Resumindo

A margem rígida é aquela que separa claramente os pontos positivos e negativos.

A margem flexível também é chamada SVM linear “barulhenta”, pois inclui alguns pontos classificados incorretamente.

Machine Learning – SVM Suport Vector Machines

Parâmetro de Regularização C



Distância Mínima Entre os VS:

$$\min_{\mathbf{w},b} \quad \frac{\|\mathbf{w}\|}{2} + C\sum_{i=1}^{N}\xi^{(i)},$$

$$\begin{aligned} s.t. \quad & y^{(i)}(\mathbf{w}^{\mathbf{T}}\mathbf{x}^{(i)} + b) \geq 1 - \xi^{(i)}, && \forall i \in \{1, \ldots, N\} \\ & \xi^{(i)} \geq 0, && \forall i \in \{1, \ldots, N\} \end{aligned}$$

Maximizar a Distância Mínima (Otimização):

$$\max_{\alpha} \quad \sum_{i=1}^{N}\alpha^{(i)} - \sum_{i=1}^{N}\sum_{j=1}^{N}\left(y^{(i)}\alpha^{(i)}\phi\left(\mathbf{x}^{(i)}\right)^{T}\phi\left(\mathbf{x}^{(j)}\right)y^{(j)}\alpha^{(j)}\right)$$

$$s.t. \quad 0 \leq \alpha^{(i)} \leq C,$$

Machine Learning – SVM Suport Vector Machines

Parâmetro de Regularização C

O parâmetro de regularização C no Modelo SVM é responsável pelo treinamento do modelo com hiperplano de margem flexível ou rígida.

Quanto maior o valor de C menor a margem do hiperplano selecionada para o treinamento de um modelo.

Quanto menor o valor de C maior a margem do hiperplano escolhida para o treinamento de um modelo.

Para obter resultados de classificação mais precisos (menos amostras classificadas incorretamente), é necessário selecionar C com grande valor.

OBRIGADO

FIAP

FIAP