Resumos DAA

Capítulo 1 - Conceitos e metodologias

CONCEITOS
Machine Learning(Aprendizagem Automática): Área que aborda as ferramentas e técnicas de construção de modelos que podem aprender sozinhas

- Produto final é **software**
- Autónomo e independente
- Algoritmos orientados aos dados

EX: arvores de regressão/classificação Regressões linear/logisitca

Data Science(Ciência de Dados): Área cientifica que estuda os dados e como extrair conhecimento e significado dos mesmos

- Produto final é **relatórios e apresentações**

A **Aprendizagem Simbólica** refere-se ao facto de todos os passos se basearem em representações simbólicas de leitura humana.
Desvantagem:

- Regras e o conhecimento precisa de ser codificado à mão

Uma das desvantagens principais da **Aprendizagem não simbólica** é que é como é que o sistema chegou a uma conclusão

- Não há decisões de alto risco

Aprendizagem Supervisionada é quando os casos que se usam para aprender incluem informação acerca dos resultados pretendidos.
Normalmente, são divididos em duas categorias:

- Classificação: quando os resultados são discretos (preto, branco, cinza...);
- Regressão: quando os resultados são contínuos (preço, temperatura idade,...).

Aprendizagem não Supervisionada é quando não são conhecidos resultados sobre os casos, apenas os enunciados dos problemas, tornando necessário a escolha de técnicas de aprendizagem que avaliem o funcionamento interno do sistema.
Normalmente, são divididos em duas categorias:

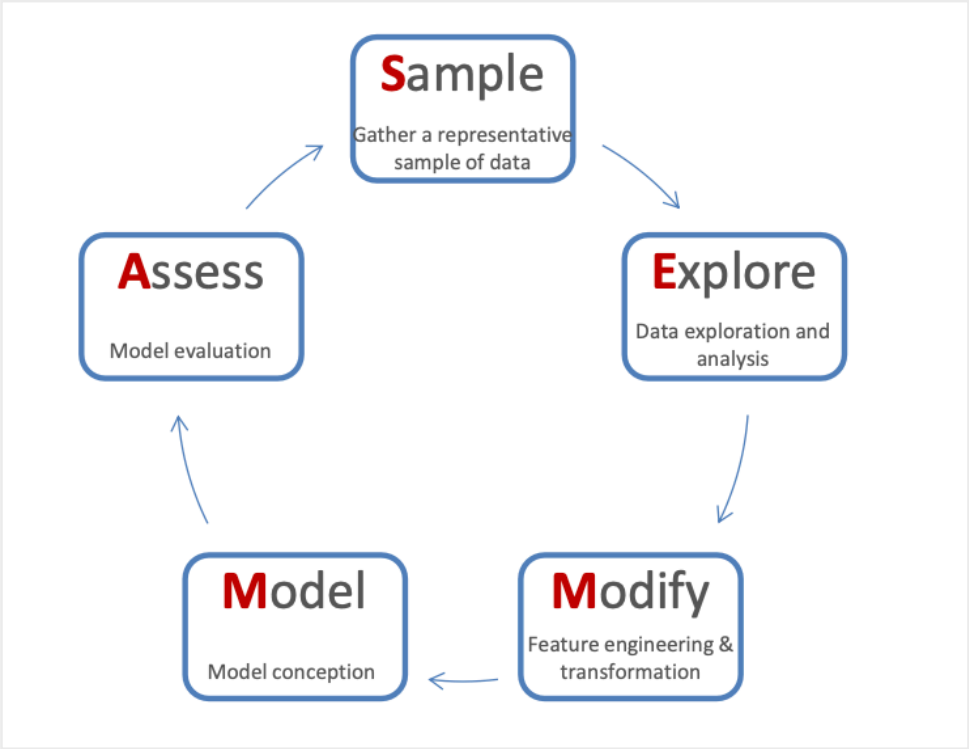
- Segmentação (Clustering): quando se pretende organizar os dados em grupos coerentes (agrupar clientes que compram produtos biológicos)
- Redução (reduction): reduzir o número de características de um conjunto de dados ou decompor o conjunto de dados em múltiplos componentes
- Associação: quando se pretende conhecer regras que associem o comportamento demonstrado pelos dados (pessoas que compram

produtos biológicos não compram produtos de charcutaria)

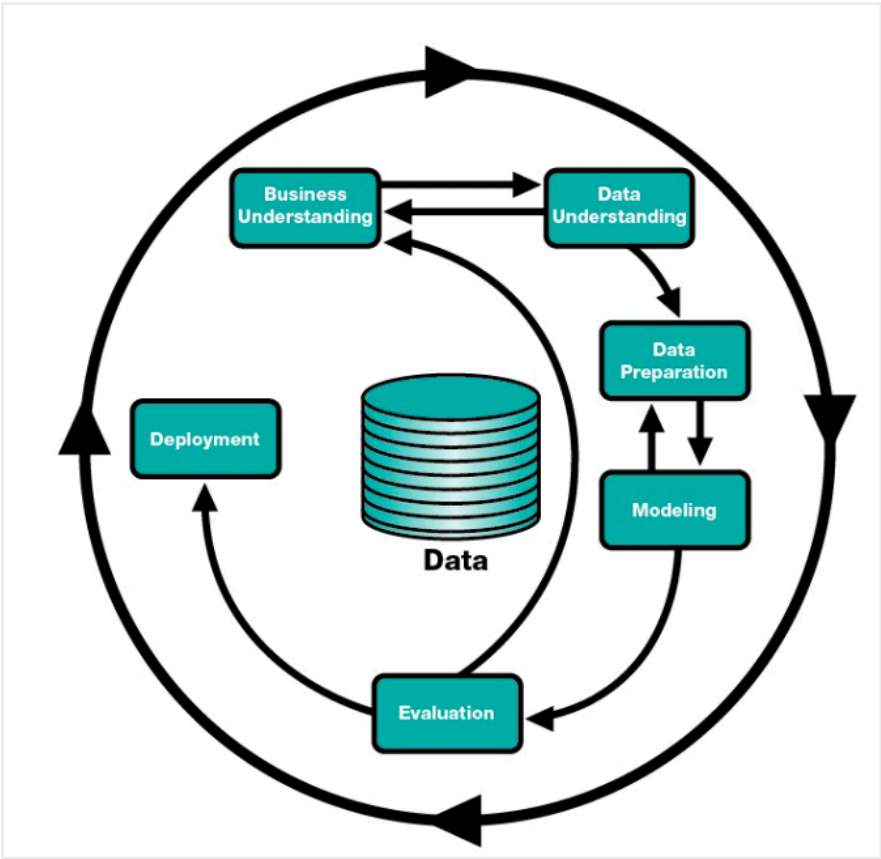
Aprendizagem por Reforço apesar de não ter informação sobre os resultados pretendidos, permite efetuar uma avaliação sobre se os resultados produzidos são bons ou maus.

METODOLOGIAS

SEMMA : Divide o processo em cinco etapas: amostrar os dados (Sample), explorar padrões e variáveis (Explore), modificar e preparar os dados (Modify), criar modelos (Model) e avaliar o desempenho (Assess). Foca na modelagem e preparação técnica dos dados



CRISP-DM: Divide o processo em seis fases: compreensão do negócio, entendimento dos dados, preparação, modelação, avaliação e desenvolvimento. Abrange desde o alinhamento com objetivos do negócio até a entrega da solução, sendo flexível e amplamente usado em diferentes setores.



NOTA:

- SEMMA foca mais na fase de modelagem e análise de dados.
- CRISP-DM é mais voltado ao ciclo de vida completo do projeto

Capítulo 2 - Exploração e Preparação dos Dados

QUALIDADE DOS DADOS - problemas

- Missing values
- Dados duplicados
- Barulho
- Outliers

EXPLORAÇÃO DOS DADOS

- Tendência Central: média, moda, mediana...
- Dispersão Estatística: variância, desvio padrão...
- Distribuição de Probabilidade: Uniforme, Exponencial...
- Correlação/Dependência: entre pares de variáveis, com a variável dependente...
- Visualização de Dados: tabelas, gráficos, boxplots...

Input Features/Input Vector (variável independente)
Target/Class/Label (variável dependente)

PREPARAÇÃO DOS DADOS - básicas
Conjunto básico de técnicas de preparação de dados:

- União e interseção de colunas
- Concatenação
- Filtros de colunas, linhas...
- Agregações básicas (média, soma...)

PREPARAÇÃO DOS DADOS - avançadas

- **Feature Scaling** - normaliza o intervalo das variáveis independentes (Normalização, ajusta a um intervalo ou Standartização, centra os valores numa média)
- **Detecção de Outliers**
- **Feature Selection** - apenas as mais importantes
- **Missing Values Treatment** - remover, colocar média, interpolar...
- **Nominal Value Discretization**
- **Binning** - valores contínuos são agrupados em intervalos, **retira informação ao modelo**
- **Feature Engineering** exemplo das datas (com apenas 29/08/2003 sabemos também que é Terça-Feira Feriado....)

Capítulo 3 - Aprendizagem Supervisionada, Regressão Linear e Logística

MODELOS LINEARES
Usados tanto para **classificação** (separação entre classes) quanto para **regressão** (previsão de valores contínuos). No entanto, eles **não resolvem problemas não lineares**, ou seja, não são eficazes quando a relação entre as variáveis não segue um padrão linear.

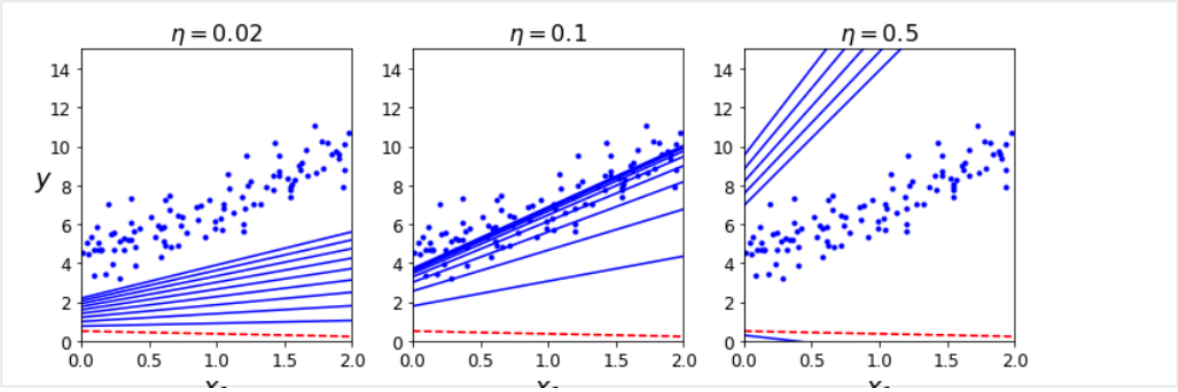
REGRESSÃO LINEAR
A **regressão linear** é uma ferramenta simples e eficaz para prever um valor contínuo com base em uma ou mais variáveis preditoras.
Ela **ajusta uma linha reta** aos dados e usa essa linha para **fazer previsões** de valores desconhecidos.

REGRESSÃO LOGÍSTICA
Usada para modelos de classificação **binária** (variável dependente categórica). Pode ser aplicada para mais que duas classes (Maçã banana e pera -> Ver se a imagem é uma maçã)

O **Least Squares Method (Método dos Mínimos Quadrados)** é uma técnica matemática usada para encontrar os **parâmetros** (coeficientes) de um modelo de regressão linear, minimizando o erro entre os valores previstos pelo modelo e os valores observados.

O **Gradient Descent (Descida do Gradiente)** é um método iterativo para otimizar parâmetros de um modelo. É usado frequentemente em regressão linear para estimar os coeficientes β0 e β1 de maneira eficiente.

Se o coeficiente for muito pequeno demora a convergir, se for muito alto pode divergir



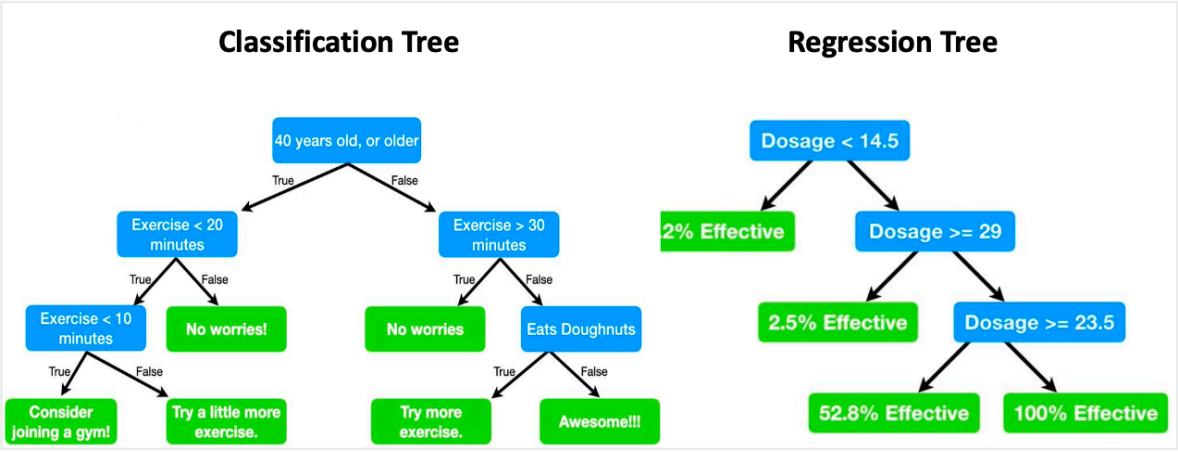
SOLUÇÕES PARA O OVERFITTING

- Reduzir o número dos atributos usados

Capítulo 4 - Árvores de Decisão

Medições de impureza:

- Gini Impurity
- Entropy Impurity
- Information Gain



Pruning (poda) é uma técnica utilizada em **árvores de decisão** para **reduzir o tamanho da árvore** e evitar o **overfitting**

A poda envolve dois processos principais:

1. **Pré-poda (Pre-Pruning)**: Interromper o crescimento da árvore antes que ela se torne excessivamente complexa.
2. **Pós-poda (Post-Pruning)**: A árvore é simplificada, removendo alguns

de seus ramos ou nós(método popularmente usado -> Cost-Complexity Pruning)

Vantagens:

- Não requer normalização dos dados
- Não requer escalonamento dos dados
- Missing values não afeta o processo de construção da árvore

Desvantagens:

- Inadequado para problemas caracterizados por muitas interações entre atributos
- Pequenas mudanças nos dados, podem provocar grandes mudanças na estrutura da árvore

Capítulo 5 - Aprendizagem em Conjunto (Ensemble Learning)

Conjuntos de modelos trabalham combinando múltiplos modelos base para contribuir um melhor

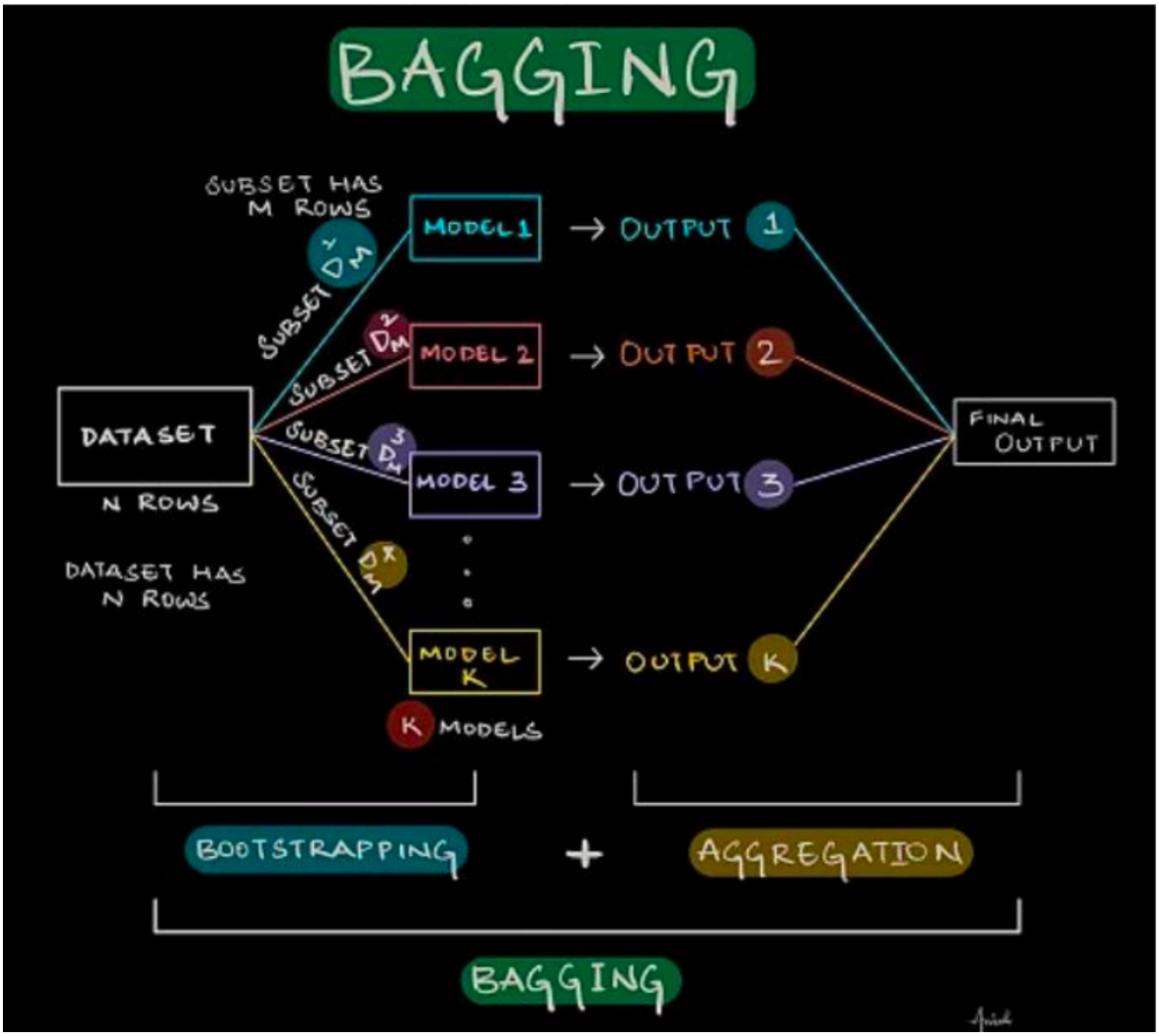
MAX VOTING

Cada modelo de machine learning faz um voto, a opção que tiver mais é selecionada. No caso do soft é escolhida a média.

Hard Voting	**Soft Voting**
Classifier_1 predicts *Class* ***A***	*Classifier_1* predicts *Class* ***A*** with the probability of 99%
Classifier_2 predicts *Class* ***B***	*Classifier_2* predicts *Class* ***A*** with the probability of 49%
Classifier_3 predicts *Class* ***B***	*Classifier_3* predicts *Class* ***A*** with the probability of 49%
2/3 classifier models predict class **B**.	The average probability of belonging to Class A is (99 + 49 + 49) / 3 = 65.67%.
Class B is the **ensemble decision**.	Thus, ***Class A*** is the **ensemble decision**.

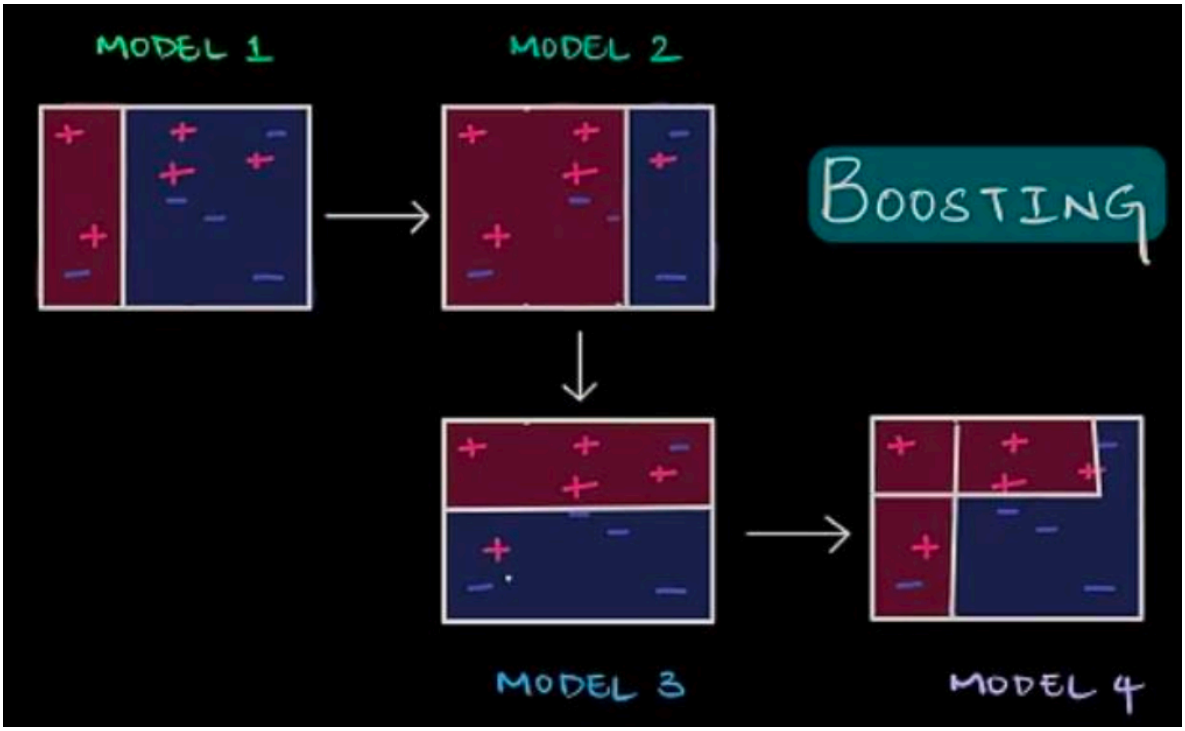
BAGGING

Utiliza modelos independentes, treinados com amostras de dados diferentes(**bootstrapping**)(como a escolha é aleatória, pode haver dados do dataset que não são usados), para melhorar a **precisão**, **estabilidade** e **robustez** do modelo final. Ele é especialmente eficaz em modelos com **alta variância**, como as **árvores de decisão**, e é amplamente utilizado em algoritmos como o **Random Forest**. Embora ofereça benefícios, o bagging também pode ser computacionalmente caro e reduzir a interpretabilidade do modelo



BOOSTING

Treina os modelos sequencialmente, ajustando-os para corrigir os erros cometidos pelos modelos anteriores



AdaBost - atribui pesos aos exemplos mal classificados para serem corrigidos no modelo anterior
XBBoost - otimiza o gradiente boosting através do gradient descent

STACKING
Os **resultados dos modelos base** são usados como **entradas** para um **meta-modelo** (ou modelo de nível superior), que aprende a combinar essas previsões

BLENDING
Idêntico ao stacking, no entanto, o blending adota uma abordagem menos rigorosa para dividir os dados e ajustar o meta-modelo

Capítulo 6 - Redes neuronais artificiais

São modelos de aprendizado de máquina inspirados no funcionamento do cérebro humano. O conhecimento é guardado nas conexões entre os neurónios.
O conhecimento é adquirido a partir de **um ambiente (dados)**, através de um **processo de aprendizagem** (algoritmo de treinamento) que ajusta os **parâmetros da rede**.

NEURONIOS ARTIFICIAIS

Os neurônios recebem entradas (inputs): Os inputs podem ser dados de características ou sinais provenientes de outros neurônios.
Pesos são associados a cada conexão: Cada conexão entre neurônios possui um peso que indica a importância relativa do input associado.
O neurônio calcula um valor de ativação: O cálculo é feito combinando os

valores de entrada ponderados pelos pesos.
O sinal filtrado é enviado como output: A saída do neurônio pode servir como input para outros neurônios ou como a previsão final da rede.

ARQUITETURA DAS REDES (TOPOLOGIA)

Há duas categorias, **supervisionado** e **não supervisionado.**

A topologia **feedforward** refere-se a uma configuração de redes neuronais em que os dados fluem em uma única direção: **da camada de entrada para a camada de saída**, passando por uma ou mais camadas ocultas

Computation of the Output:

- **Forward Pass**:
  - Os dados percorrem a rede da entrada até a saída.
- **Saída Final**:
  - Os resultados das camadas ocultas são combinados na camada de saída, gerando o valor final da rede (e.g., probabilidades de classes ou valores contínuos).

Training: Ajustar os valores dos pesos das conexões que minimizam a perda função: no caso de RNAs, generalização da função de custo de regressão logística

Backpropagation

Aspecto	Descrição
Base do Método	Utiliza o gradiente da função de erro para ajustar os pesos (semelhante ao gradiente descendente).
Taxa de Aprendizado (η)	Define o tamanho dos passos na direção oposta ao gradiente.
Épocas	Número de vezes que o conjunto de treinamento é usado para atualizar os pesos.
Mini-Batches	Dados divididos em subconjuntos menores para processamento mais eficiente.
Inicialização	Pesos inicializados aleatoriamente.
Critérios de Parada	Número fixo de épocas, limite de tempo ou convergência com base em dados de validação.

Fases :

- **Forward Pass**:
  - Calcula a saída da rede com base nos pesos atuais.
  - Mede o erro ao comparar as previsões com os valores reais.

- **Backward Pass**:
    - Calcula os gradientes do erro em relação aos pesos usando a regra da cadeia.

Learning curves são gráficos que ajudam a avaliar o desempenho de um modelo durante o processo de treinamento. Eles mostram como a **função de custo** ou o **erro** muda em relação ao número de épocas ou ao tamanho do conjunto de dados. Essas curvas são uma ferramenta essencial para diagnosticar problemas no treinamento e ajustar o modelo.

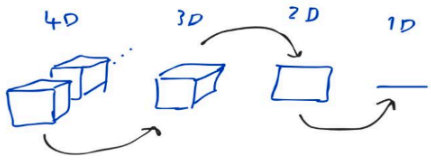
VANTAGENS:
- Interpretabilidade
- Capacidade de modelar relações não lineares

DESVANTAGENS:
- Overfitting
- Demora no treino

Capítulo 7 - Não supervisionado

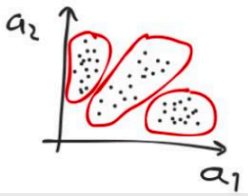
Tarefas

Dimensionality Reduction - task of reducing the number of input features in a dataset (not samples).



Anomaly Detection - task of detecting instances that are very different from the norm.

Clustering - task of grouping similar instances into clusters.



4

ALGORITMOS

Dimensionality Reduction:
Principal Component Analysis (**PCA**);
Manifold Learning - LLE, Isomap, **t-SNE**;
Autoencoders and others.

Anomaly Detection:
Isolation Forest;
Local Outlier Factor;
Minimum Covariance Determinant;
other algorithms initially designed for dimensionality reduction or supervised learning.

Clustering:
K-Means;
Hierarchical Clustering and Spectral Clustering;
DBSCAN and OPTICS;
Affinity Propagation;
Mean Shift and BIRCH;
Gaussian Mixture Models;

Anomaly detection é a tarefa de identificar instâncias ou padrões **anormais ou incomuns** dentro de um conjunto de dados.

Clustering é uma tarefa de aprendizado não supervisionado que tem como objetivo agrupar pontos de dados não rotulados em **clusters** (ou grupos) de forma que os pontos de dados dentro de um mesmo cluster sejam mais **semelhantes entre si** do que com os pontos de dados de outros clusters

- **K-means** - Baseado em **partição**
- **DBSCAM** - Baseado em **densidade**, o que significa que ele agrupa pontos de dados que estão próximos uns dos outros com base em uma medida de densidade.

Capítulo 8 - Support Vector Machine

É um dos algoritmos mais poderosos e populares usados em **aprendizado supervisionado**, tanto para **classificação** quanto para **regressão.**
O principal objetivo do SVM é encontrar uma **hiperplano** que separa os dados de diferentes classes da melhor forma possível

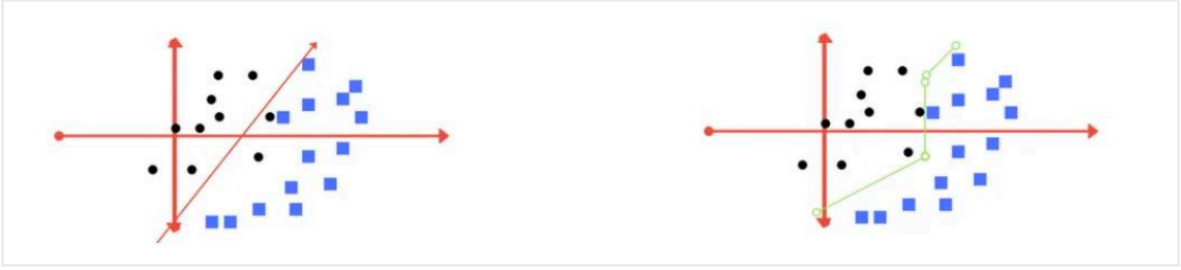
Tuning parameters: Kernel, Regularization and Gamma.

Em muitos casos, os dados não são linearmente separáveis. Para lidar com isso, o SVM utiliza o **truque do kernel**. Isso envolve a transformação dos dados originais para um espaço de **dimensão superior** onde os dados se tornam linearmente separáveis

Regularização (C) controla a **penalização** das violações das margens:

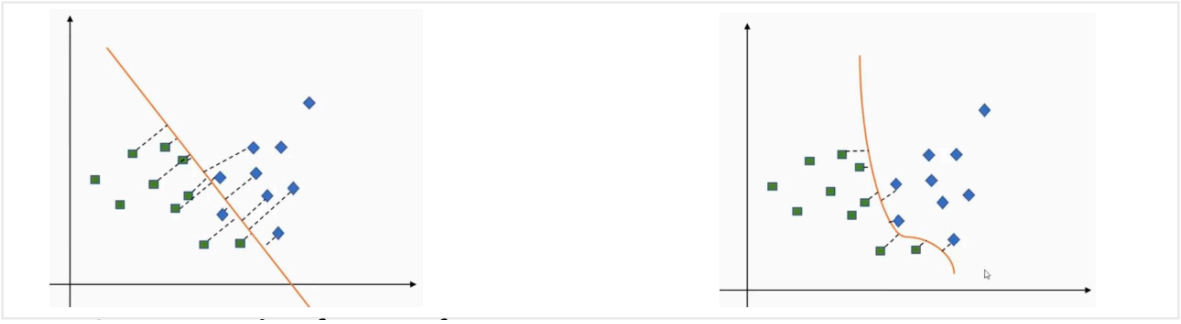
- **C baixo**: mais flexível, mais tolerante a erros, mas pode causar underfitting.
- **C alto**: menos tolerante a erros, mais ajustado aos dados, mas pode

causar overfitting.



Gamma controla a **influência de cada ponto de dado** sobre a fronteira de decisão:

- **Gamma baixo**: maior influência de cada ponto, fronteira de decisão mais suave, pode causar underfitting.
- **Gamma alto**: influência local, fronteira de decisão mais complexa e ajustada, pode causar overfitting.



Capítulo 9 - Aprendizagem por reforço (Reinforcement Learning)

Reinforcement learning é definido pelas características de aprendizagem do problema. Programamos as characteristics do problema para aprender;
É o estado atual que decide o que fazer no futuro

Aspecto	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Tipo de Dados	Dados rotulados (entrada e saída)	Dados não rotulados (apenas entrada)	Sem rótulos, o agente interage com o ambiente
Objetivo	Prever saídas a partir de entradas	Descobrir padrões ou estruturas nos dados	Maximizar recompensa por meio de ações em um ambiente
Exemplos	Classificação, Regressão	Clustering, Redução de Dimensionalidade	Jogos, Robótica, Sistemas de Recomendação
Algoritmos Comuns	Regressão Linear, SVM, Árvores de Decisão	K-Means, PCA, Algoritmos de Agrupamento	Q-Learning, Deep Q-Networks, Algoritmos de Política
Vantagens	Resultados previsíveis com dados rotulados	Não precisa de rótulos, pode descobrir padrões ocultos	Pode aprender com feedback contínuo e em tempo real
Desvantagens	Requer dados rotulados, pode ser sensível a ruído	Difícil de avaliar, sem "resposta certa"	Computacionalmente caro, aprendizado pode ser demorado

A **exploration** refere-se ao comportamento do agente de **tentar novas ações** que ele ainda não experimentou no ambiente

A **exploitation** envolve o agente escolher **ações com base no conhecimento adquirido** durante o treino

Comparação entre Q-Learning e SARSA

Característica	Q-Learning	SARSA
Tipo de Política	Off-policy (não depende da política seguida)	On-policy (depende da política seguida)
Fórmula de Atualização	Usa $\max_{a'} Q(s', a')$	Usa $Q(s', a')$, ação escolhida no próximo estado
Exploração vs Exploração	Pode explorar mais agressivamente	Segue mais estritamente a política atual
Ajuste da Política	A política ótima pode ser aprendida mais rapidamente	A política precisa ser aprimorada durante o treinamento

Q-Learning (off-policy): Atualiza Q(s, a) usando o valor max{a'}Q(s',a'), o que significa que o agente aprende a política ótima, independentemente das ações exploratórias tomadas durante a exploração.
SARSA (on-policy): Atualiza Q(s,a) usando Q(s′,a′), que é a ação realmente tomada no estado s′, refletindo a política que o agente está seguindo no momento (que pode ser exploratória ou otimizada)