

Machine Learning

Conceito, Classificação, Exemplos

Jean Daniel Dias

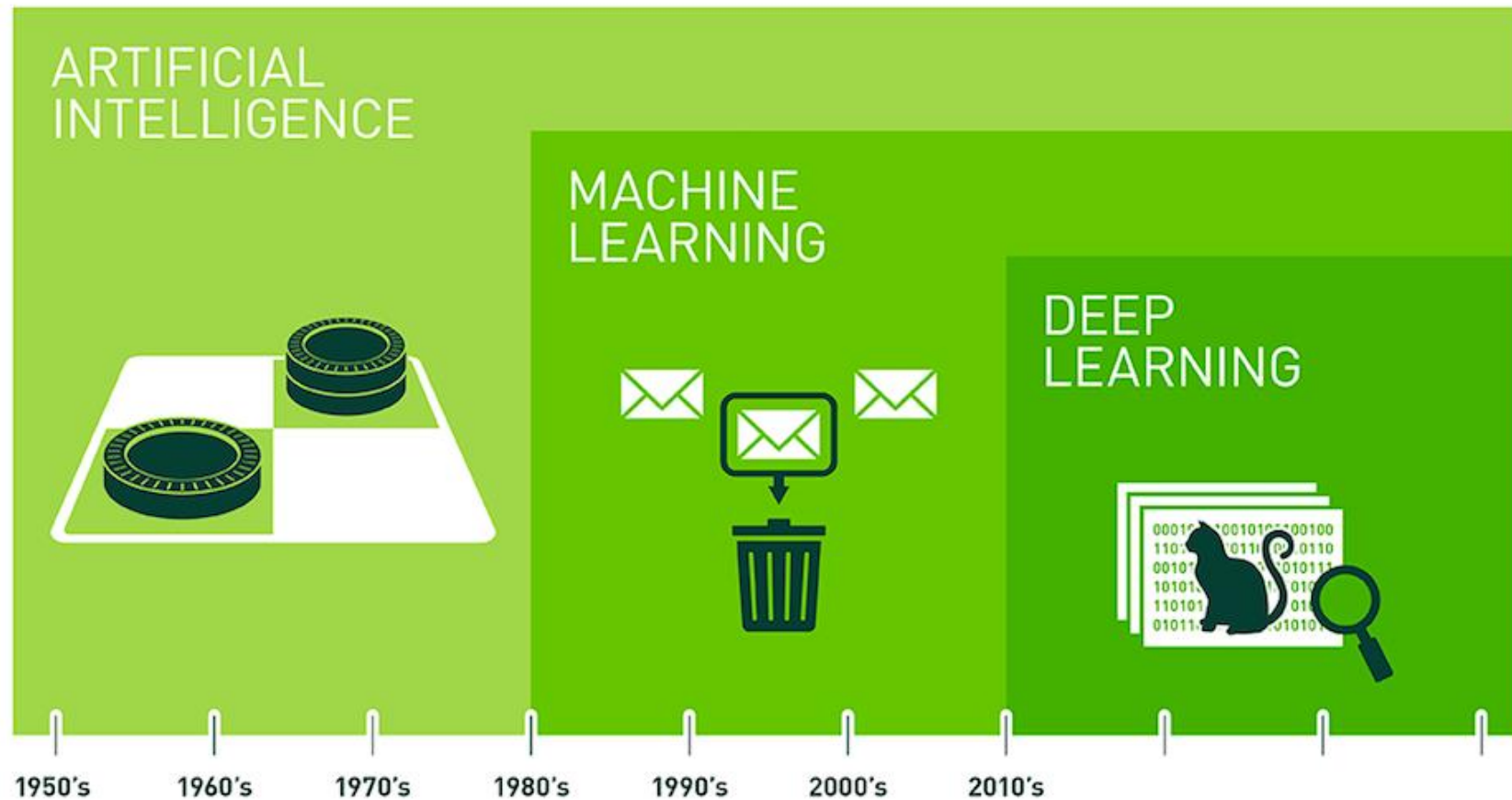
João Victor Magalhães

Pedro Henrique Lopes de Carvalho

Machine Learning – Considerações Iniciais

- *Artificial Intelligence, Machine Learning e Deep Learning* são as três grandes divisões que se fazem na área de Inteligência Artificial;
- São concêntricas entre si: a A.I. que veio primeiro, engloba as demais;
- Deep Learning é área inscrita em M.L. que permitiu o *boom* da expansão de A.I. nos últimos anos.

Machine Learning – Considerações Iniciais



Machine Learning – O que é?

- *Machine Learning* da maneira mais básica é a prática de usar algoritmos para coletar dados, aprender com eles, e então fazer uma determinação ou predição sobre alguma coisa no mundo;
- Conjunto específico de instruções para completar uma tarefa em particular, a máquina é “treinada” usando uma quantidade grande de dados e algoritmos que dão a ela a habilidade de aprender como executar a tarefa;
- No Brasil é comum utilizar “Aprendizado de Máquina” em substituição ao termo em inglês.

Machine Learning – Como funciona?

- Quando se desenvolve um sistema de aprendizado de máquina, a estrutura utilizada na programação é diferente da programação de software tradicional. No **método tradicional** se cria um conjunto de regras para gerar uma resposta a partir do processamento dos dados introduzidos;
- Já os **algoritmos de *Machine Learning*** são criados a partir dos dados que serão analisados e as as repostas (ou resultados) que se esperam dessa análise, no final do processo o sistema cria as próprias regras ou perguntas.

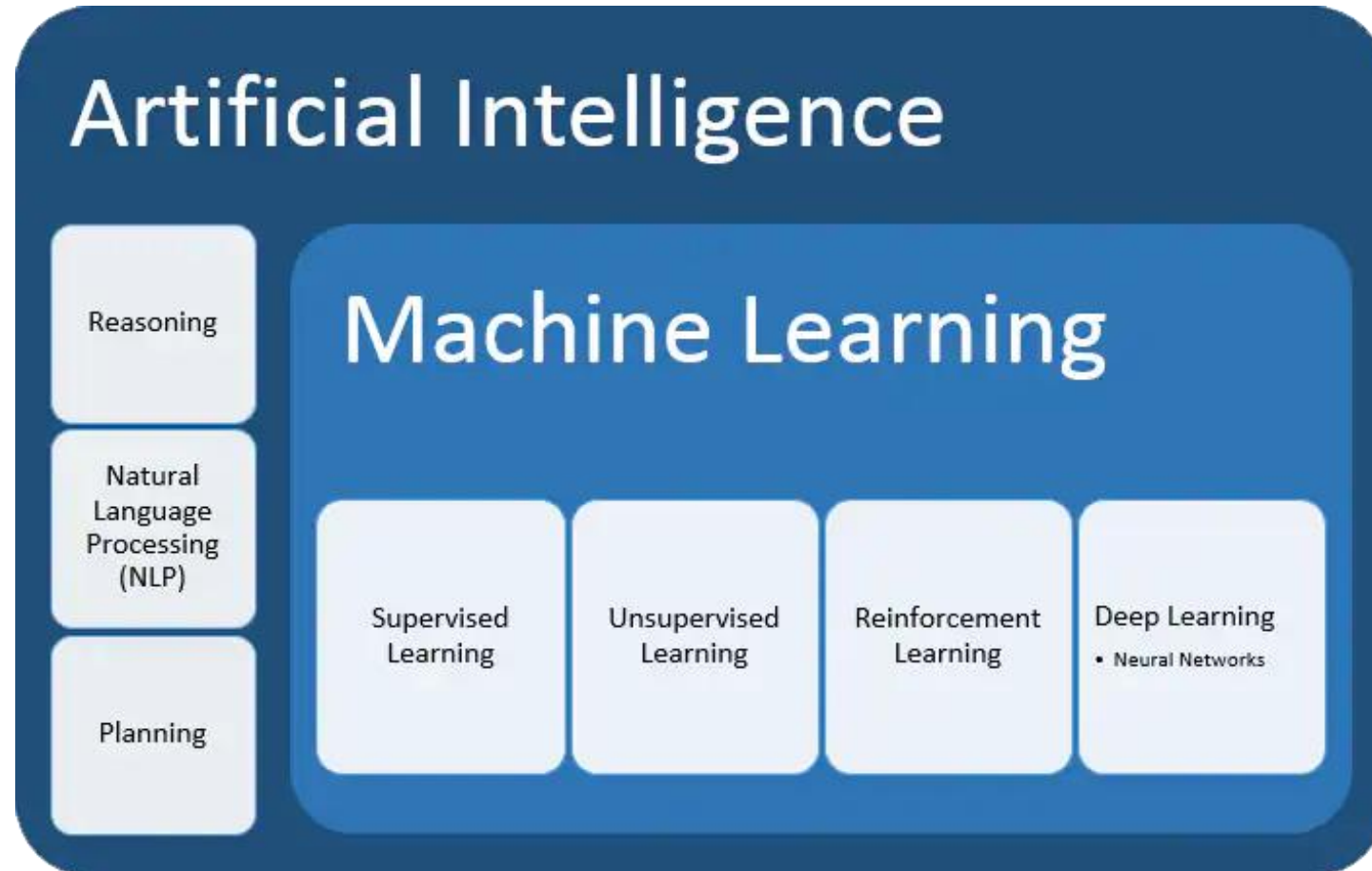
Machine Learning e Big Data

- Para trabalhar com o sistema de aprendizado de máquina é necessário utilizar um certo conjunto de dados.
O Big Data permite que os dados sejam virtualizados para que possam ser armazenados da maneira mais eficiente e econômica, seja on premises ou na cloud. Além da eficiência o Big Data também auxilia na melhoria da velocidade e confiabilidade da rede, removendo outras limitações físicas associadas ao gerenciamento de dados em grande quantidade.
- Apesar das vantagens oferecidas no processo, uma empresa não necessita ter Big Data para trabalhar com Machine Learning.

Machine Learning - Classificação

- Dividido em três classificações:
 - Supervisionado
 - Não Supervisionado
 - Aprendizado por Reforço
- Há ainda a *Deep Learning*, comumente classificada em duas:
 - Redes Neurais Profundas
 - Processamento de Linguagem Natural

Machine Learning - Classificação



Machine Learning - Supervisionado

- O aprendizado supervisionado geralmente começa com um conjunto estabelecido de dados e um certo entendimento de como esses dados são classificados. O aprendizado supervisionado destina-se a encontrar padrões em dados que possam ser aplicados em um processo analítico. Esses dados rotularam recursos que definem o significado dos dados.
- Por exemplo, é possível criar um aplicativo de *machine learning* que distinga entre milhões de animais, com base em imagens e descrições escritas.

Machine Learning – Não Supervisionado

- O aprendizado não supervisionado é usado quando o problema requer uma grande quantia de dados não rotulados. Por exemplo, aplicativos de mídia social, como Twitter, Instagram e Snapchat, têm grandes quantias de dados não rotulados. Entender o significado por trás desses dados requer algoritmos que classificam os dados com base nos padrões ou clusters encontrados.
- O aprendizado não supervisionado conduz um processo iterativo, analisando dados sem intervenção humana. É usado com tecnologia de detecção de spam por e-mail, por exemplo.

Machine Learning – Aprendizado por Reforço

- O aprendizado por reforço é um modelo de aprendizado comportamental. O algoritmo recebe feedback da análise de dados, orientando o usuário para o melhor resultado. O aprendizado de reforço difere de outros tipos de aprendizado supervisionado, porque o sistema não é treinado com o conjunto de dados de amostra. Em vez disso, o sistema aprende por meio de tentativa e erro. Portanto, uma sequência de decisões bem-sucedidas resultará no processo sendo reforçado, porque resolve melhor o problema em questão.

Machine Learning – Deep Learning:

Redes Neurais

- *Deep learning* é um método específico de aprendizado de máquina que incorpora redes neurais em camadas sucessivas para aprender com os dados de uma maneira iterativa. *Deep learning* é especialmente útil quando você está tentando aprender padrões de dados não estruturados.
- Redes neurais complexas de *Deep learning* são projetadas para emular como o cérebro humano funciona, para que os computadores possam ser treinados para lidar com abstrações e problemas mal definidos. Redes neurais e *deep learning* são frequentemente usados em aplicativos de reconhecimento de imagem, fala e visão computacional.

Aprendizado Supervisionado: K-nearest neighbors e linear models

- O KNN é um algoritmo não paramétrico, onde a estrutura do modelo será determinada pelo dataset utilizado.
- Este algoritmo também é conhecido como de aprendizado lento ou melhor dizendo, é um algoritmo preguiçoso, o termo certo é “lazy”. Os algoritmos do tipo lazy, não necessitam de dados de treinamento para se gerar o modelo, o que diminui em partes o processo inicial, mas em contrapartida gerará uma necessidade de análise posterior mais apurada.
- No caso de algoritmos que não necessitam de treinamento, todos os dados obtidos no dataset serão utilizados na fase de teste, resultando em um treinamento muito rápido e em um teste e validação lentos, momento o qual necessitamos estar bem atentos aos resultados gerados.

Aprendizado Supervisionado: K-nearest neighbors e linear models

- Vantagens:

- ✓ É um dos mais simples algoritmos a se implementar.
- ✓ Boa precisão na maioria dos casos aplicados.
- ✓ A facilidade para se efetuar tuning dos poucos parâmetros existentes (Valor de K e Medida de Distância)
- ✓ Tempo para se efetuar treinamento dos dados também é diferenciado, sendo um dos mais rápidos para esta atividade.
- ✓ O KNN pode ser utilizado em dados não lineares bem como para problemas de regressão.

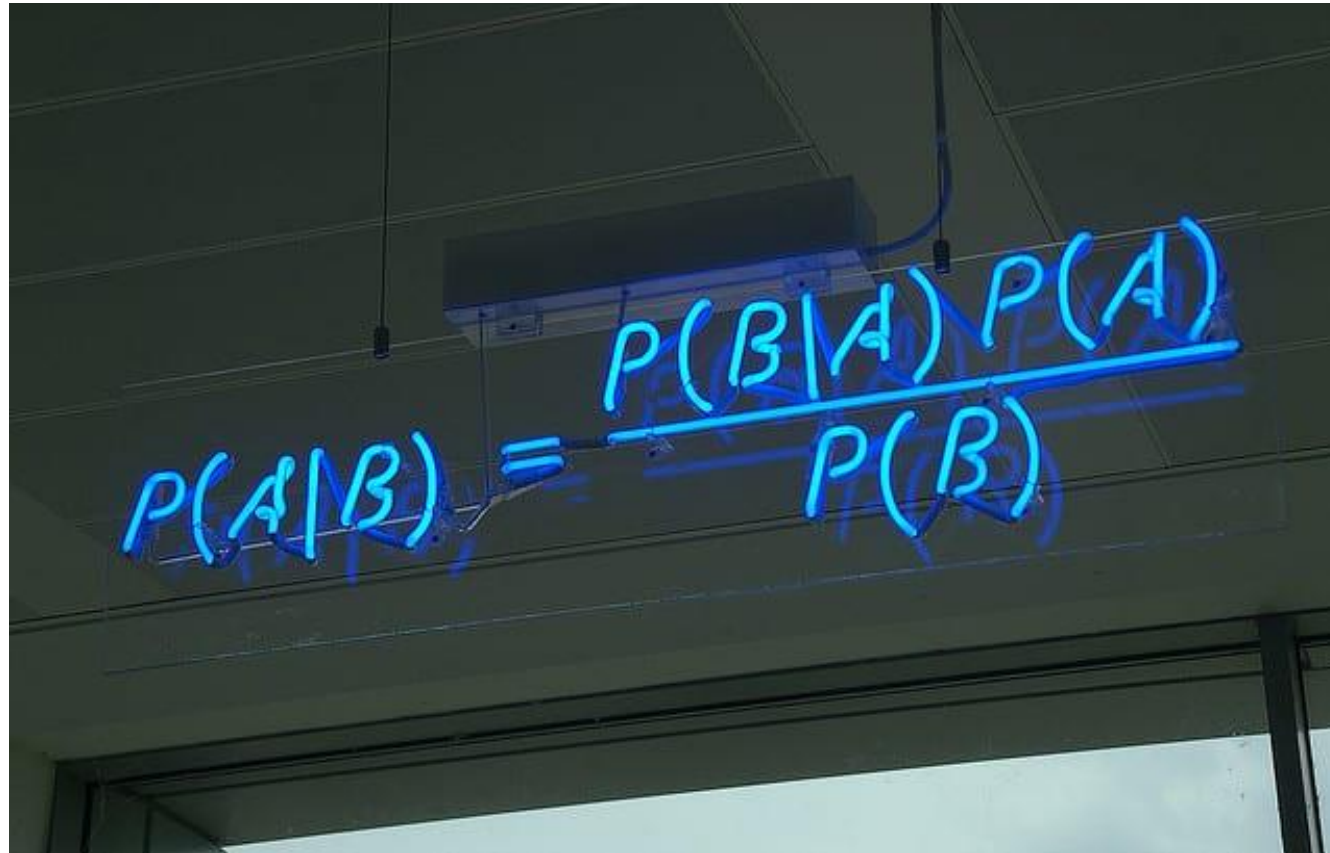
Aprendizado Supervisionado: K-nearest neighbors e linear models

- Desvantagens:
 - Uma demora excessiva na fase de teste e o alto consumo de memória para realizar esta atividade de teste, uma vez que o mesmo armazena todo dataset em memória.
 - KNN não é indicado a dados de grandes dimensões, imagina ter que calcular a distância de todos os pontos de dados entre si, quanto mais dados e mais dimensões desses registros, maior será o tempo para processar os cálculos.
 - Uma vez que o mesmo trabalha com medidas de distância necessita sempre que nos atentemos a escala dos valores utilizados, para que não gere resultados equivocados.

Aprendizado Supervisionado: Naïve bayes classifiers

- O algoritmo “Naive Bayes” é um classificador probabilístico baseado no “Teorema de Bayes”, o qual foi criado por Thomas Bayes (1701 - 1761) para tentar provar a existência de Deus.
- Atualmente, o algoritmo se tornou popular na área de Aprendizado de Máquina (*Machine Learning*) para categorizar textos baseado na frequência das palavras usadas, e assim pode ser usado para identificar se determinado e-mail é um SPAM ou sobre qual assunto se refere determinado texto, por exemplo.

Aprendizado Supervisionado: Naïve bayes classifiers



A photograph of a blue neon sign mounted on a dark ceiling. The sign displays the Naïve Bayes formula for calculating the probability of class A given evidence B. The formula is written in a stylized, handwritten font. The numerator consists of the conditional probability $P(B|A)$ multiplied by the prior probability $P(A)$. The denominator is the marginal probability $P(B)$. The entire expression is enclosed in a blue neon frame.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Aprendizado Supervisionado: Naïve bayes classifiers

- **Vantagens:**

- ✓ É fácil e rápido para prever o conjunto de dados da classe de teste. Também tem um bom desempenho na previsão de classes múltiplas.
- ✓ Quando a suposição de independência prevalece, um classificador Naive Bayes tem melhor desempenho em comparação com outros modelos como regressão logística, e você precisa de menos dados de treinamento.
- ✓ O desempenho é bom em caso de variáveis categóricas de entrada comparada com a variáveis numéricas. Para variáveis numéricas, assume-se a distribuição normal (curva de sino, que é uma suposição forte).

Aprendizado Supervisionado: Naïve bayes classifiers

- Desvantagens:
 - Se a variável categórica tem uma categoria (no conjunto de dados de teste) que não foi observada no conjunto de dados de treinamento, então o modelo irá atribuir uma probabilidade de 0 (zero) e não será capaz de fazer uma previsão. Isso é muitas vezes conhecido como “Zero Frequency”. Para resolver isso, podemos usar a técnica de alisamento. Uma das técnicas mais simples de alisamento é a chamada estimativa de Laplace.
 - Por outro lado naive Bayes é também conhecido como um mau estimador, por isso, as probabilidades calculadas não devem ser levadas muito a sério.
 - Outra limitação do Naive Bayes é a suposição de preditores independentes. Na vida real, é quase impossível que ter um conjunto de indicadores que sejam completamente independentes.

Aprendizado Supervisionado: Decision trees e ensembles of decisions trees

- Árvore de decisão é um tipo de algoritmo de aprendizagem supervisionada (com uma variável alvo pré-definida), muito utilizada em problemas de classificação.
- Ele funciona para ambas as variáveis categóricas e contínuas de entrada e de saída. Na árvore de decisão, dividimos a população ou amostra em dois ou mais conjuntos homogêneos (ou sub-populações) com base nos divisores/diferenciadores mais significativos das variáveis de entrada.

Aprendizado Supervisionado: Decision trees e ensembles of decisions trees

$$Entropia(S) = \sum p_i \log_2 p_i$$

$$Ganho(S, A) = Entropia(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} \cdot Entropia(S_v)$$

- S é o conjunto de exemplo de treino;
- p_+ é a porção de exemplos positivos;
- p_- é a porção de exemplos negativos;
- A entropia é dada pelo desdobramento da equação 1

Aprendizado Supervisionado: Decision trees e ensembles of decisions trees

- **Vantagens**

- ✓ **Fácil de entender:** A visualização de uma árvore de decisão torna o problema fácil de compreender, mesmo para pessoas que não tenham perfil analítico. Não requer nenhum conhecimento estatístico para ler e interpretar. Sua representação gráfica é muito intuitiva e permite relacionar as hipóteses também facilmente.
- ✓ **Útil em exploração de dados:** A árvore de decisão é uma das formas mais rápidas de identificar as variáveis mais significativas e a relação entre duas ou mais variáveis. Com a ajuda de árvores de decisão, podemos criar novas variáveis/características que tenham melhores condições de prever a variável alvo.
- ✓ **Menor necessidade de limpar dados:** Requer menos limpeza de dados em comparação com outras técnicas de modelagem. Até um certo nível, não é influenciado por pontos fora da curva “outliers” nem por valores faltantes (“missing values”).
- ✓ **Não é restrito por tipos de dados:** Pode manipular variáveis numéricas e categóricas.
- ✓ **Método não paramétrico:** A árvore de decisão é considerada um método não-paramétrico. Isto significa que as árvores de decisão não pressupõe a distribuição do espaço nem a estrutura do classificador.

Aprendizado Supervisionado: Decision trees e ensembles of decisions trees

- **Desvantagens**

- **Sobreajuste (“Over fitting”)**: Sobreajuste é uma das maiores dificuldades para os modelos de árvores de decisão. Este problema é resolvido através da definição de restrições sobre os parâmetros do modelo e da poda (discutido em mais detalhes abaixo).
- **Não adequado para variáveis contínuas**: ao trabalhar com variáveis numéricas contínuas, a árvore de decisão perde informações quando categoriza variáveis em diferentes categorias

Aprendizado Supervisionado: Support vector machines (SVM)

- Support Vector Machine” (SVM) é um [algoritmo de aprendizado de máquina supervisionado](#) que pode ser usado para desafios de classificação ou regressão. No entanto, é usado principalmente em problemas de classificação.
- Nesse algoritmo, plotamos cada item de dados como um ponto no espaço n-dimensional (onde n é o número de recursos que você tem), com o valor de cada recurso sendo o valor de uma determinada coordenada. Então, nós executamos a classificação encontrando o hiperplano que diferencia muito bem as duas classes
- Vetores de Suporte são simplesmente as coordenadas da observação individual. Support Vector Machine é uma fronteira que melhor segrega as duas classes (hiperplano / linha).

Aprendizado Supervisionado: Support vector machines (SVM)

$$f(x) = wx + b = 0$$

- Onde w é o vetor de pesos (mesma dimensão das amostras) perpendicular ao hiperplano de separação e b é um escalar.

Aprendizado Supervisionado: Support vector machines (SVM)

- Vantagens

- ✓ Funciona muito bem com margem de separação clara
- ✓ É eficaz em espaços dimensionais elevados.
- ✓ É eficaz nos casos em que o número de dimensões é maior que o número de amostras.
- ✓ Ele usa um subconjunto de pontos de treinamento na função de decisão (chamados de vetores de suporte), portanto, também é eficiente em termos de memória

Aprendizado Supervisionado: Support vector machines (SVM)

- Desvantagens
 - Não funciona bem quando temos um grande conjunto de dados porque o tempo de treinamento necessário é maior
 - Também não funciona muito bem, quando o conjunto de dados tem mais ruído, ou seja, as classes de destino estão sobrepostas
 - O SVM não fornece estimativas de probabilidade diretamente, elas são calculadas usando uma valiosa [validação cruzada](#) de cinco vezes. É o método SVC relacionado da biblioteca scikit-learn do Python.

Aprendizado Não Supervisionado:

Análise de componentes principais (PCA)

- PCA calcula a projeção dos dados em algum vetor que maximize a variância dos dados e perca a menor quantidade de informação possível. Surpreendentemente, estes vetores são os [autovetores](#) da matriz de correlação das características de um conjunto de dados..
- PCA é usado em uma variedade de casos de uso:
 1. Visualize dados multidimensionais . As visualizações de dados são uma ótima ferramenta para comunicar dados multidimensionais como gráficos bidimensionais ou tridimensionais.
 2. Comprimir informações . A Análise de Componentes Principais é usada para compactar informações para armazenar e transmitir dados com mais eficiência. Por exemplo, pode ser usado para [comprimir imagens](#) sem perder muita qualidade, ou no processamento do sinal. A técnica foi aplicada com sucesso em uma ampla gama de problemas de compressão em reconhecimento de padrões (especificamente reconhecimento de rosto), reconhecimento de imagem e muito mais.

Aprendizado Não Supervisionado:

Análise de componentes principais (PCA)

3. Simplifique decisões de negócios complexas . O PCA foi empregado para simplificar decisões de negócios tradicionalmente complexas. Por exemplo, os traders usam mais de 300 instrumentos financeiros para gerenciar carteiras. O algoritmo tem se mostrado bem-sucedido no [gerenciamento de risco de carteiras de derivativos de taxa de juros](#) , reduzindo o número de instrumentos financeiros de mais de 300 para apenas 3-4 componentes principais.

4. Esclareça processos científicos complicados . O algoritmo tem sido aplicado extensivamente na compreensão de fatores complicados e multidirecionais, que aumentam a probabilidade de conjuntos neurais [dispararem potenciais de ação](#) .

Quando o PCA é usado como parte do pré-processamento, o algoritmo é aplicado a:

1. Reduza o número de dimensões no conjunto de dados de treinamento.
2. Remova o ruído dos dados. Como o PCA é calculado localizando os componentes que explicam a maior quantidade de variação, ele captura o sinal nos dados e omite o ruído.

Aprendizado Não Supervisionado:

Análise de componentes principais (PCA)

- Vantagens:
 - ✓ Fácil de calcular . O PCA é baseado em álgebra linear, que é computacionalmente fácil de resolver por computadores.
 - ✓ Acelera outros algoritmos de aprendizado de máquina . Os algoritmos de aprendizado de máquina convergem mais rápido quando treinados nos componentes principais em vez do conjunto de dados original.
 - ✓ Neutraliza os problemas de dados de alta dimensão . Dados de alta dimensão fazem com que algoritmos baseados em regressão se ajustem facilmente. Ao usar o PCA de antemão para diminuir as dimensões do conjunto de dados de treinamento, evitamos que os algoritmos preditivos sejam superdimensionados.

Aprendizado Não Supervisionado:

Análise de componentes principais (PCA)

- Desvantagens:

1. Baixa interpretabilidade dos componentes principais . Os componentes principais são combinações lineares dos recursos dos dados originais, mas não são tão fáceis de interpretar. Por exemplo, é difícil dizer quais são os recursos mais importantes no conjunto de dados após calcular os componentes principais.
2. O trade-off entre perda de informação e redução de dimensionalidade . Embora a redução de dimensionalidade seja útil, ela tem um custo. A perda de informações é uma parte necessária do PCA. Equilibrar a compensação entre redução de dimensionalidade e perda de informações é, infelizmente, um compromisso necessário que devemos fazer ao usar o PCA.

Aprendizado Não Supervisionado: K-means clustering

- Este algoritmo é capaz de realizar o treinamento de um modelo para fazer o agrupamento de objetos, ele trabalha com processos de similaridade, ou seja, a principal ideia é encontrar itens semelhantes um com os outros, e mais distintos possíveis dentre os membros de outros clusters/grupos de acordo com seus atributos.
- Você tem uma base de dados com um total de n dados, $X = \{x_1, x_2, \dots, x_n\}$ cada ponto x com quantas dimensões você quiser.
- Você quer agrupa-los em k clusters, $C = \{C_1, C_2, \dots, C_k\}$
- A inércia de um cluster C_i pode ser calculada por:

Aprendizado Não Supervisionado: K-means clustering

$$\sum_{x \in C_i} \|x - \mu_i\|^2$$

- i é o índice do cluster que você está analisando, e μ_i portanto é a média do cluster i , que é também o centro do cluster i . $\|x - \mu_i\|$ é a distância de um ponto x ao centro do cluster i . \sum significa soma, e $x \in C_i$ quer dizer para os x que pertencem ao cluster C_i . Ou seja, lendo tudo junto: para os pontos x que estão no cluster C_i , some o quadrado das distâncias de cada ponto até o centro desse cluster.

Aprendizado Não Supervisionado:K-means clustering

- Vantagem:

- ✓ menos sensível a outliers

- Desvantagem:

- Maior complexidade computacional devido a etapa de ordenação

Aprendizado Não Supervisionado:

Agrupamento hierárquico

- É uma técnica de clusterização de dados que baseia-se no tamanho e distância dos dados em um conjunto.
- Quando temos uma massa de dados tendemos a agrupa-las visualmente pela proximidade ou pela cor, criando pequenos grupos que se diferenciam quando analisados por uma determinada característica, essa técnica chama-se clustering,

Aprendizado Não Supervisionado: Agrupamento hierárquico

- Distância euclidiana

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$$

Fórmula da distância euclidiana

- Distância euclidiana quadrada

$$\|a - b\|_2^2 = \sum_i (a_i - b_i)^2$$

Fórmula da distância euclidiana quadrada

- Distância manhattan

$$\|a - b\|_1 = \sum_i |a_i - b_i|$$

Fórmula da distância manhattan

- Distância máxima

$$\|a - b\|_\infty = \max_i |a_i - b_i|$$

Fórmula da distância máxima

Aprendizado Não Supervisionado: Agrupamento hierárquico

- **Vantagens:**

- ✓ Dados subjacentes tiverem uma estrutura hierárquica (como as correlações nos mercados financeiros) e você desejar recuperar a hierarquia. Você ainda pode aplicar os meios para fazer isso, mas pode acabar com partições (do mais grosseiro (todos os pontos de dados de um cluster) até o mais fino (cada ponto de dados é um cluster)) que não está aninhado e, portanto, não é uma hierarquia adequada.
- ✓ Se você deseja se aprofundar nas propriedades mais refinadas do cluster, talvez não queira opor o cluster simples, como -eans, ao cluster hierárquico, como os Links Unico, Médio e Completo. Por exemplo, todos esses clusters economizam espaço, ou seja, quando você está construindo clusters, não distorce o espaço, enquanto um cluster hierárquico como Ward não economiza espaço, ou seja, a cada etapa da fusão, distorce o espaço métrico.

Aprendizado Não Supervisionado:

Agrupamento hierárquico

- **Desvantagens:**

- Para concluir, as desvantagens dos algoritmos hierárquicos de clustering podem ser muito diferentes entre si. Alguns podem compartilhar propriedades semelhantes a k-means: Ward visa otimizar a variação, mas o Single Linkage não. Mas eles também podem ter propriedades diferentes: Ward é dilatador de espaço, enquanto o Single Linkage é conservador de espaço, como k-médias.

Aprendizado Não Supervisionado:

DBSCAN

- O algoritmo DBSCAN é um método de agrupamento baseado em densidade. Permite identificar grupos de diferentes formatos topológicos e também descartar ruídos isolados.
- O DBSCAN normalmente apresenta bons resultados, no entanto, ele realiza diversos cálculos de distâncias no processo de agrupamento, possuindo baixa eficiência em grandes conjuntos de dados. Neste trabalho é apresentado um novo método amostral que possibilita aplicar o DBSCAN em um conjunto reduzido de exemplos, aproximando-se do resultado original do algoritmo aplicado sobre todo o conjunto de dados.
- Portanto, o método torna possível a execução de grandes conjuntos de dados com resultados semelhantes ao do DBSCAN na base de dados completa, porém, com maior eficiência.

Aprendizado Não Supervisionado:

DBSCAN

- Vantagens:
 - ✓ DBSCAN does not require one to specify the number of clusters in the data a priori, as opposed to [k-means](#).
 - ✓ DBSCAN can find arbitrarily-shaped clusters. It can even find a cluster completely surrounded by (but not connected to) a different cluster. Due to the MinPts parameter, the so-called single-link effect (different clusters being connected by a thin line of points) is reduced.
 - ✓ DBSCAN has a notion of noise, and is robust to [outliers](#).
 - ✓ DBSCAN requires just two parameters and is mostly insensitive to the ordering of the points in the database. (However, points sitting on the edge of two different clusters might swap cluster membership if the ordering of the points is changed, and the cluster assignment is unique only up to isomorphism.)
 - ✓ DBSCAN is designed for use with databases that can accelerate region queries, e.g. using an [R* tree](#).
 - ✓ The parameters minPts and ϵ can be set by a domain expert, if the data is well understood.

Aprendizado Não Supervisionado:

DBSCAN

- Desvantagens:
 1. DBSCAN is not entirely deterministic: border points that are reachable from more than one cluster can be part of either cluster, depending on the order the data are processed. For most data sets and domains, this situation does not arise often and has little impact on the clustering result:^[4] both on core points and noise points, DBSCAN is deterministic. DBSCAN*^[8] is a variation that treats border points as noise, and this way achieves a fully deterministic result as well as a more consistent statistical interpretation of density-connected components.
 2. The quality of DBSCAN depends on the [distance measure](#) used in the function `regionQuery(P,ε)`. The most common distance metric used is [Euclidean distance](#). Especially for [high-dimensional data](#), this metric can be rendered almost useless due to the so-called "[Curse of dimensionality](#)", making it difficult to find an appropriate value for ϵ . This effect, however, is also present in any other algorithm based on Euclidean distance.
 3. DBSCAN cannot cluster data sets well with large differences in densities, since the `minPts-ε` combination cannot then be chosen appropriately for all clusters.^[9]
 4. If the data and scale are not well understood, choosing a meaningful distance threshold ϵ can be difficult

Deep Learning - Redes Neurais:

Processamento de Linguagem Natural (PLN)

- O processamento de linguagem natural incorpora técnicas diversas para interpretar a linguagem humana, desde métodos estatísticos e de machine learning a abordagens algorítmicas e baseadas em regras. Nós precisamos de uma boa variedade de abordagens, porque dados baseados em texto ou voz divergem muito, assim como suas aplicações práticas.
- Knapp (1982) resumiu brevemente os casos em que a busca em LN pode obter melhor desempenho: para tópicos específicos; para temas atuais; para novas terminologias ainda não incluídas nas LC; para uma busca retrospectiva em que o conceito da LC é muito recente e não cobre os anos anteriores; quando o termo da LC é muito abrangente ou muito específico; para pesquisa em várias bases de dados; para identificação imediata de palavras de títulos dos documentos; para complementação de citação bibliográfica incompleta.

Deep Learning - Redes Neurais: Processamento de Linguagem Natural (PLN)

- Vantagens:
 - ✓ Permite o imediato registro da informação em uma base de dados, sem necessidade de consulta a uma linguagem de controle.
 - ✓ Processo de busca é facilitado com a ausência de treinamentos específicos no uso de uma linguagem de controle.
 - ✓ Termos de entrada de dados são extraídos diretamente dos documentos que vão constituir a base de dados.
 - ✓ Temas específicos citados nos documentos podem ser encontrados.
 - ✓ Elimina os conflitos de comunicação entre os indexadores e os usuários, pois ambos terão acesso aos mesmos termos.

Deep Learning - Redes Neurais:

Processamento de Linguagem Natural (PLN)

- Desvantagens:
 - Os usuários da informação, no processo de busca, precisam fazer um esforço intelectual maior para identificar os sinônimos, as grafias alternativas, os homônimos etc. Haverá alta incidência de respostas negativas ou de relações incorretas entre os termos usados na busca (por ausência de padronização).
 - Custos de acesso tendem a aumentar com a entrada de termos de busca aleatórios. Uma estratégia de busca que arrole todos os principais conceitos e seus sinônimos deve ser elaborada para cada base de dados (ex: nomes comerciais de substâncias químicas não ocorrem no Chemical Abstracts).
 - Perda de confiança do usuário em uma possível resposta negativa.

Referências

- <https://medium.com/data-science-brigade/a-diferen%C3%A7a-entre-intelig%C3%A7%C3%A2ncia-artificial-machine-learning-e-deep-learning-930b5cc2aa42>
- <https://www.ibm.com/br-pt/analytics/machine-learning>
- *Machine Learning for Dummies*, disponível em: <https://www.ibm.com/common/ssi/cgibin/ssialias?htmlfid=IMM14209USEN>
- <https://www.deeplearningbook.com.br/>