

Aula de Nivelamento

Introdução a Algoritmos

Instalação da linguagem e pacotes

As aulas técnicas deste curso serão todas feitas utilizando a linguagem Python em sua versão 3.5+.

- [Google Colab](#)
- [Python](#)
- [Anaconda](#)

Operações lógicas

AND, *NAND*, *OR*, *XOR* e *NOT*.

AND		
A	B	Saída
0	0	0
0	1	0
1	0	0
1	1	1

NAND		
A	B	Saída
0	0	1
0	1	1
1	0	1
1	1	0

OR		
A	B	Saída
0	0	0
0	1	1
1	0	1
1	1	1

XOR		
A	B	Saída
0	0	0
0	1	1
1	0	1
1	1	0

NOT	
A	Saída
0	1
1	0

Falando computação

Como então conversar com a máquina?

```
var1 ← Verdadeiro
```

```
var2 ← Falso
```

```
Se var1 e var2, então faça:
```

```
    imprime "Primeira linha."
```

```
ou se var1 ou var2, então faça:
```

```
    imprime "Segunda linha."
```

```
senão, faça:
```

```
    imprime "Terceira linha."
```

Tipos de dados

Tipo	Descrição	Sintaxe
String	Uma cadeia de caracteres	"Olá mundo!"
Integer	Número inteiro	1; 2; 43; 9000
Float	Número com ponto flutuante	1.2; 3.1416; 9.6
Boolean	Valores booleanos	True; False; 0; 1
Complex	Números complexos	4+5i
Dict	Dicionário	{ 'chave' : valor }
List	Lista mutável	[1, 2, 3, 4, 5]
Tuple	Tupla imutável	(1, 2, 3, 4, 5)

Operadores aritméticos

Operador	Descrição
+	Adição
-	Subtração
*	Multiplicação
/	Divisão
%	Módulo (resto)
//	Divisão (parte inteira)
**	Exponenciação

Operadores básicos de comparação

Operador	Descrição	Sintaxe
Igualdade	Compara a igualdade entre dois valores	==
Maior que	Verifica se o primeiro valor é maior que o segundo valor	>
Menor que	Verifica se o primeiro valor é menor que o segundo valor	<
Maior ou igual	Verifica se o primeiro valor é maior ou igual ao segundo valor	>=
Menor ou igual	Verifica se o primeiro valor é menor ou igual ao segundo valor	<=
Diferente	Verifica se o primeiro valor é diferente do segundo valor	!=

Condicionais

Utilizam lógicas booleanas para decidir qual bloco de instrução satisfaz a comparação.

```
Se condição, então:
```

```
    Faça isso.
```

```
Ou se condição, então:
```

```
    Faça isso.
```

```
Senão:
```

```
    Faça isso.
```

Loops

Varrem uma lista/array de informação e executam um bloco de ação para cada item encontrado ou para cada instância de execução.

```
Para cada item em uma lista de itens, faça:  
    Bloco de ação...
```

Funções

Funções são blocos de código capazes de efetuar uma determinada ação toda vez em que são chamados. Podem ou não receber valores como parâmetros de entrada.

```
def nome_da_funcao(variavel):  
    # Bloco de código dentro do escopo da função
```

Probabilidade e Inferência Estatística

Conceitos gerais

- **Descritiva**
 - Organiza, demonstra e resume dados..

Conceitos gerais

- **Descritiva**
 - Organiza, demonstra e resume dados.
- **Probabilidade**
 - Analisa situações sujeitas ao acaso..

Conceitos gerais

- **Descritiva**

- Organiza, demonstra e resume dados.

- **Probabilidade**

- Analisa situações sujeitas ao acaso.

- **Inferência**

- Obter respostas sobre um fenômeno com dados representativos.

Métodos estatísticos

- **Observação**

- Os elementos estudados, não são afetados. Sua função aqui é a de observador.
 - **Ex.** Você realiza uma pesquisa de satisfação com os pacientes do pronto atendimento em relação ao tempo de atendimento.

Métodos estatísticos

- **Observação**

- Os elementos estudados, não são afetados. Sua função aqui é a de observador.
 - **Ex.** Você realiza uma pesquisa de satisfação com os pacientes do pronto atendimento em relação ao tempo de atendimento.

- **Experimento**

- Você impõe ao elementos suas condições e busca uma relação de causa/efeito.
 - **Ex.** Realiza uma pesquisa de eficiência de um novo tratamento.
 - Grupo X
 - Grupo Y
 - Grupo de controle (placebo)

Variáveis

- **Qualitativas** - Expressam uma categoria
 - **Nominais**: Categorias sem hierarquia (Ex. Masculino e Feminino);
 - **Ordinais**: Categorias com hierarquia (Ex. Júnior, Pleno e Sênior).

Variáveis

- **Qualitativas** - Expressam uma categoria

- **Nominais:** Categorias sem hierarquia (Ex. Masculino e Feminino);
- **Ordinais:** Categorias com hierarquia (Ex. Júnior, Pleno e Sênior).

- **Quantitativas** - Numerais

- **Contínuas:** Valores reais, podem assumir qualquer intervalo (Ex. 35000; 1.53; - 0.003);
- **Discretas:** Números fixos, inteiros e num intervalo (Ex. [1; 2; 3; ...]; 10; [10.20; 10.40; 10.60]).

Amostra

Entendemos como **amostra** uma parcela de alguma **população** extraída utilizando uma **técnica específica** que garanta chance igual para todos os indivíduos desta em serem selecionados na extração.

Se provado que não existe **aleatoriedade** na escolha de elementos, dizemos que a amostra é **enviesada**, ou seja, não é justa.

Amostra - Conceitos

- **População**
 - São todos os elementos que deseja-se estudar.

Amostra - Conceitos

- **População**

- São todos os elementos que deseja-se estudar.

- **Amostra**

- Subconjunto dos elementos que deseja-se estudar.

Amostra - Conceitos

- **População**

- São todos os elementos que deseja-se estudar.

- **Amostra**

- Subconjunto dos elementos que deseja-se estudar.

- **Censo**

- Pesquisa com toda a população. Não apenas com elementos que deseja-se estudar.

Amostra - Conceitos

- **População**

- São todos os elementos que deseja-se estudar.

- **Amostra**

- Subconjunto dos elementos que deseja-se estudar.

- **Censo**

- Pesquisa com toda a população. Não apenas com elementos que deseja-se estudar.

- **Enviesamento**

- Não foi retirada corretamente.

Amostra - Conceitos

- **População**

- São todos os elementos que deseja-se estudar.

- **Amostra**

- Subconjunto dos elementos que deseja-se estudar.

- **Censo**

- Pesquisa com toda a população. Não apenas com elementos que deseja-se estudar.

- **Enviesamento**

- Não foi retirada corretamente.

- **Benefícios em utilizar amostras**

- Realizar um censo é custoso e demorado;
- Uma amostra apresenta as mesmas características da população que deseja-se estudar.

Amostra - Tipos

- **Aleatória simples**

- Você define um tamanho da amostra e determina a forma com que utilizará da aleatoriedade para a extração desta amostra da população (Com reposição, sem reposição).

Amostra - Tipos

- **Aleatória simples**

- Você define um tamanho da amostra e determina a forma com que utilizará da aleatoriedade para a extração desta amostra da população (Com reposição, sem reposição).

- **Estratificada**

- As amostras são selecionadas através de um estrato específico (um status específico), como por exemplo quem possui convênio médico e quem não tem.

Amostra - Tipos

- **Aleatória simples**

- Você define um tamanho da amostra e determina a forma com que utilizará da aleatoriedade para a extração desta amostra da população (Com reposição, sem reposição).

- **Estratificada**

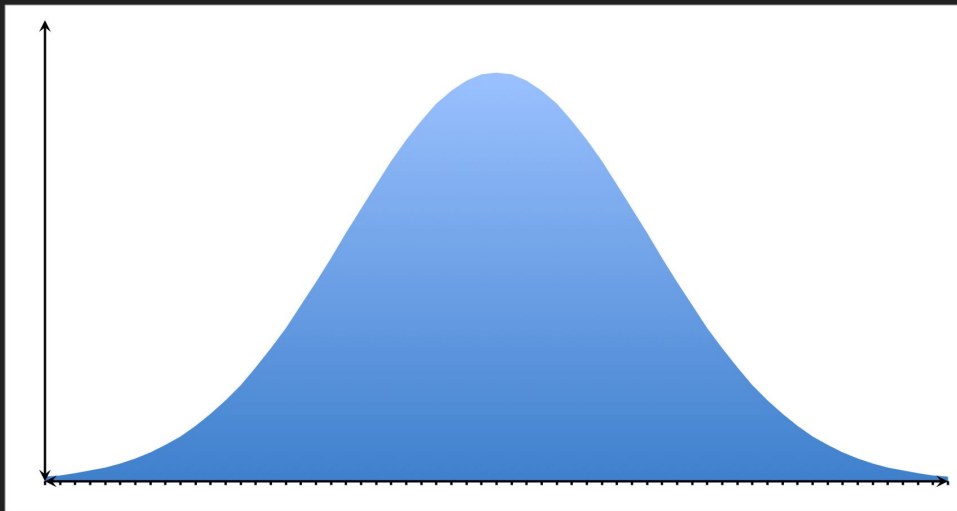
- As amostras são selecionadas através de um estrato específico (um status específico), como por exemplo quem possui convênio médico e quem não tem.

- **Sistemático**

- Existe uma progressão aritmética na escolha da próxima amostra;
- utilizado quando não se tem controle do tamanho da população;
- Quando o primeiro elemento da amostra é escolhido aleatoriamente, o próximo elemento é o enésimo próximo elemento.

Centralidade

Estudamos o centro dos dados, como eles estão centralizados.



Centralidade

- **Média**

- É a soma de todos os i valores de interesse divididos pelo total N dos valores de interesse.

Centralidade

- **Média**

- É a soma de todos os i valores de interesse divididos pelo total N dos valores de interesse.

- **Moda**

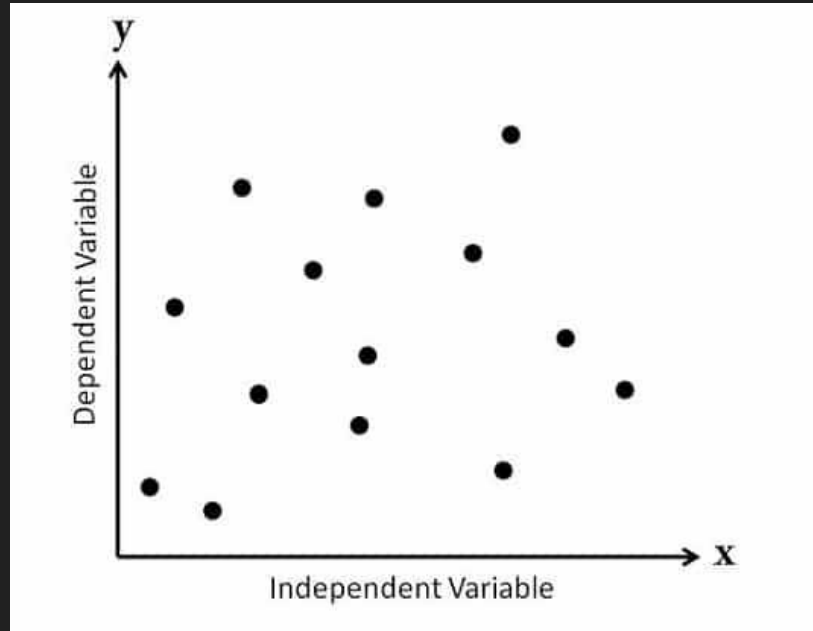
- É o valor mais frequente dentro do universo da amostra.

Centralidade

- **Média**
 - É a soma de todos os i valores de interesse divididos pelo total N dos valores de interesse.
- **Moda**
 - É o valor mais frequente dentro do universo da amostra.
- **Mediana**
 - É o valor do **meio** dentro do universo da amostra ordenada.

Variabilidade

Estudamos a distância entre os dados, como eles se diferenciam.



Variabilidade

- **Variância**
 - Mede qual a dispersão dos dados ao seu valor médio.

Variabilidade

- **Variância**

- Mede qual a dispersão dos dados ao seu valor médio.

- **Desvio padrão**

- Qual é a distância dos dados ao seu valor médio.

Variabilidade

- **Variância**

- Mede qual a dispersão dos dados ao seu valor médio.

- **Desvio padrão**

- Qual é a distância dos dados ao seu valor médio.

- **Amplitude**

- É a diferença entre o menor valor e o maior valor dos dados.

Variabilidade

- **Variância**

- Mede qual a dispersão dos dados ao seu valor médio.

- **Desvio padrão**

- Qual é a distância dos dados ao seu valor médio.

- **Amplitude**

- É a diferença entre o menor valor e o maior valor dos dados.

- **Não centrais: Quartis**

- Permite identificar variações dentro dos valores de interesse.
 - **Q1:** 25% dos menores valores;
 - **Q2:** 50%, igual a mediana;
 - **Q3:** 75% dos maiores valores.

Probabilidade

A probabilidade P é sempre um valor dentro do intervalo $0 \leq P \leq 1$.

- $P = 1$ é dito como evento certo;
 - A probabilidade de P ocorrer é 100%.
- $P = 0$ é dito como evento impossível.
 - A probabilidade de P ocorrer é 0%.

Eventos

- **Excludentes**

- Quando não podem ocorrer ao mesmo tempo.
 - **Ex.** Em um **exame de próstata** onde o paciente é do sexo **feminino**.

Eventos

- **Excludentes**

- Quando não podem ocorrer ao mesmo tempo.
 - **Ex.** Em um **exame de próstata** onde o paciente é do sexo **feminino**.

- **Não excludentes**

- Quando podem ocorrer ao mesmo tempo.
 - **Ex.** Em um **exame de próstata** onde o paciente é do sexo **masculino** e possui **convênio médico**.

Eventos

- **Excludentes**

- Quando não podem ocorrer ao mesmo tempo.
 - **Ex.** Em um **exame de próstata** onde o paciente é do sexo **feminino**.

- **Não excludentes**

- Quando podem ocorrer ao mesmo tempo.
 - **Ex.** Em um **exame de próstata** onde o paciente é do sexo **masculino** e possui **convênio médico**.

- **Dependentes**

- Um evento necessariamente precisa que outro tenha ocorrido.
 - **Ex.** Um **paciente de emergência** que tenha origem em um hospital do **SUS**.

Eventos

- **Excludentes**

- Quando não podem ocorrer ao mesmo tempo.
 - **Ex.** Em um **exame de próstata** onde o paciente é do sexo **feminino**.

- **Não excludentes**

- Quando podem ocorrer ao mesmo tempo.
 - **Ex.** Em um **exame de próstata** onde o paciente é do sexo **masculino** e possui **convênio médico**.

- **Dependentes**

- Um evento necessariamente precisa que outro tenha ocorrido.
 - **Ex.** Um **paciente de emergência** que tenha origem em um hospital do **SUS**.

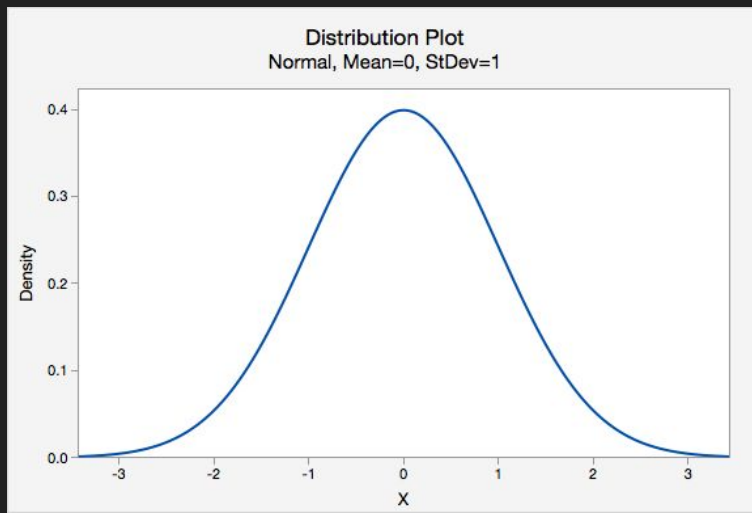
- **Independentes**

- Um evento não depende que outro tenha ocorrido.
 - **Ex.** Dois **pacientes de emergência** que um tenha origem em um hospital do **SUS** e o segundo de transferência **interna**.

Distribuição

Permite analisar o padrão da frequência dos valores de uma certa amostra, assim entendendo como seus valores são distribuídos.

Desta forma é possível inferir sobre um conjunto onde identifica-se um padrão.



Distribuição

- **Binomial**

- Permite avaliar a probabilidade da ocorrência de eventos discretos.
 - **Ex.** Na fila de um **pronto atendimento**, qual a probabilidade de chamarmos **4 vezes** um **idoso** em **7 senhas** chamadas em sequência.

Distribuição

- **Binomial**

- Permite avaliar a probabilidade da ocorrência de eventos discretos.
 - **Ex.** Na fila de um **pronto atendimento**, qual a probabilidade de chamarmos **4 vezes** um **idoso** em **7 senhas** chamadas em sequência.

- **Normal**

- A média dos dados encontra-se no centro e portanto é uma distribuição simétrica, ou seja, os valores são distribuídos da mesma forma acima e abaixo deste valor médio.
 - **Ex.** Comparar entre duas populações conhecidas **A** e **B** as probabilidades de que seus indivíduos tenham altura maior ou igual a **180 cm**.

Ciência de Dados

Ciência de dados

- Área interdisciplinar;
- Estuda e analisa dados sejam eles estruturados ou não;
- Tem como objetivo extrair informação dos dados analisados;
- Utiliza grande volume de dados junto a técnicas de aprendizado de máquina;
- Ciência de dados ❤️ estatística.

Ciência de dados

- **Dados estruturados**

- Qualquer informação **que respeita** uma estrutura ou padronização (organizado).
 - **Ex.** Uma tabela **SQL**.

Ciência de dados

- **Dados estruturados**

- Qualquer informação **que respeita** uma estrutura ou padronização (organizado).
 - **Ex.** Uma tabela **SQL**.

- **Dados não estruturados**

- Qualquer informação **que não respeita** uma estrutura ou padronização (organizado).
 - **Ex.** Uma **imagem** ou um **laudo médico**.

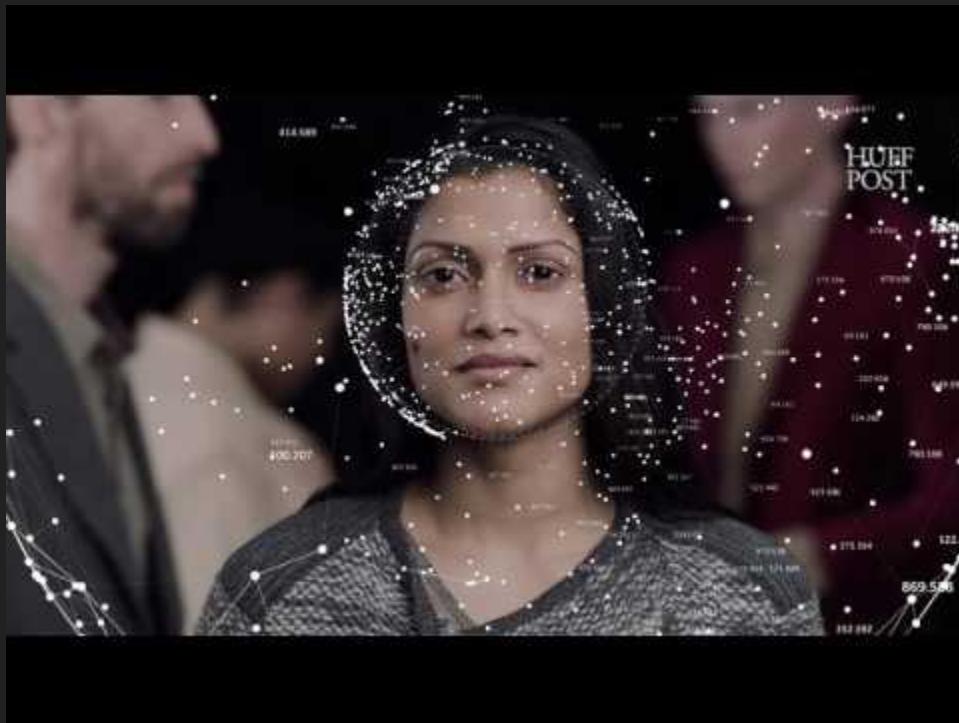
Ciência de dados

- **Importância da ciências de dados atualmente**
 - Grande volume de informação (**Big Data**);
 - Acesso a recursos para processar um grande volume de informação.

Ciência de dados

- **Aplicação de ciência de dados**
 - Ferramentas de recomendações;
 - Logística;
 - Suporte à decisão;
 - Sistemas de busca;
 - Reconhecimento de imagens.

Ciência de dados



Ciência de dados



Ciência de dados

- **Data Mining**

- Processo de exploração;
- Procura-se evidenciar padrões em um volume grande de informação;
- Utiliza-se de métodos estatísticos e algoritmos de aprendizado de máquina;
- Auxilia na descoberta de conhecimento.

Ciência de dados

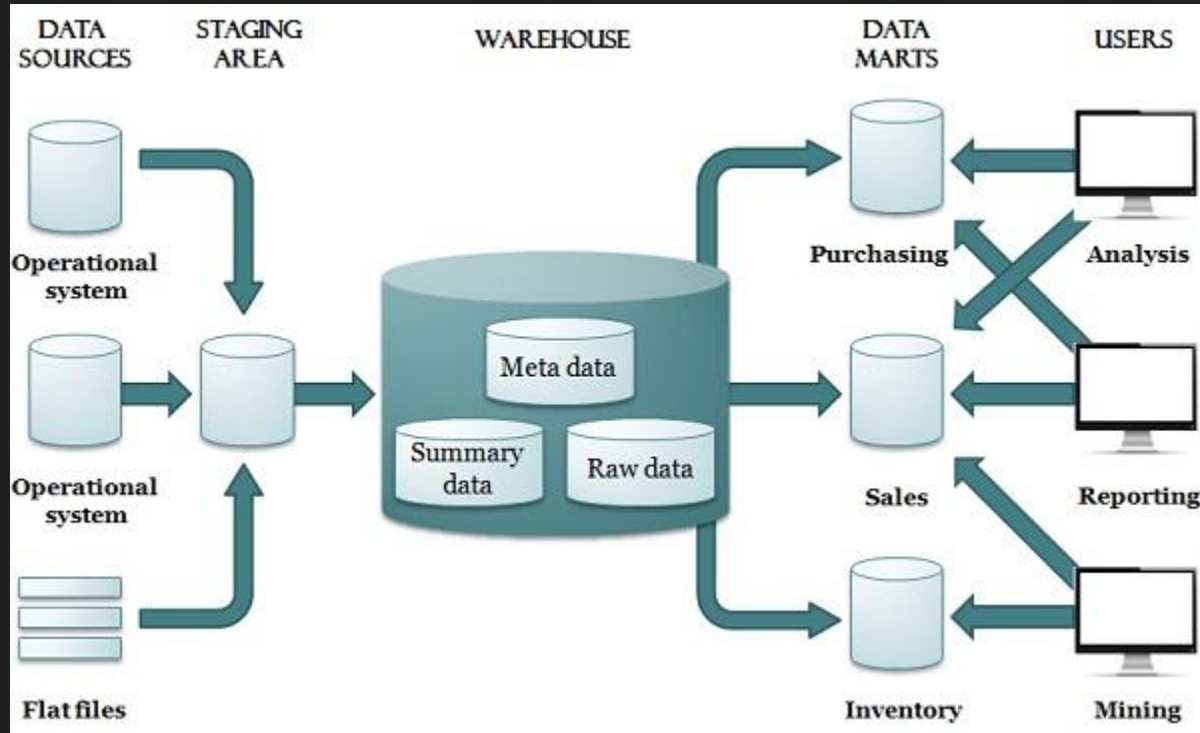
- **Data Mining**

- Processo de exploração;
- Procura-se evidenciar padrões em um volume grande de informação;
- Utiliza-se de métodos estatísticos e algoritmos de aprendizado de máquina;
- Auxilia na descoberta de conhecimento.

- **Processo de uma mineração de dados**

- Primeiro existe um processo que busca simplificar o dado que vem de uma fonte conhecida, como um relatório médico, por exemplo;
- Cria-se então os repositórios de dados (DW, Data Marts);
- Quando temos os dados segmentados podemos então selecionar nossas informações de interesse;
- É importante refinar as informações até que os padrões apareçam;
- Encontrar padrões obriga que os dados necessariamente sejam simplificados, desconsiderando aquilo que é específico e privilegiando tudo o que for generalista.

Ciência de dados



Ciência de dados



Objetivos com mineração de dados

- **Classificação**

- Permite identificar características dentro de grupos de informações que possibilitam classificar dados desconhecidos.
 - **Ex.** Paciente com certos dados de triagem, vai ou não precisar ser internado?

Objetivos com mineração de dados

- **Classificação**

- Permite identificar características dentro de grupos de informações que possibilitam classificar dados desconhecidos.
 - **Ex.** Paciente com certos dados de triagem, vai ou não precisar ser internado?

- **Clusters**

- Permite identificar agrupamentos em uma massa de dados quando ainda não foram definidos grupos.
 - Ex. De um total de exames realizados durante um mês de operação, separe os valores dos exames em 3 grupos distintos (**barato**, **médio**, **caro**).

Objetivos com mineração de dados

- **Classificação**

- Permite identificar características dentro de grupos de informações que possibilitam classificar dados desconhecidos.
 - **Ex.** Paciente com certos dados de triagem, vai ou não precisar ser internado?

- **Clusters**

- Permite identificar agrupamentos em uma massa de dados quando ainda não foram definidos grupos.
 - Ex. De um total de exames realizados durante um mês de operação, separe os valores dos exames em 3 grupos distintos (**barato**, **médio**, **caro**).

- **Associações**

- permite identificar ocorrências diretamente ligadas a um evento específico.
 - Ex. Em um dia normal, o hospital aplica em média 20 injeções de um remédio para gripe. No inverno essa média triplica.

Objetivos com mineração de dados

- **Sequência**

- Permite identificar ocorrências ao longo do tempo
 - **Ex.** Em um caso de internação, 80% das vezes existe necessidade de prescrever soro fisiológico, e que em 20% das vezes, também existe necessidade de prescrever um medicamento para febre.

Objetivos com mineração de dados

- **Sequência**

- Permite identificar ocorrências ao longo do tempo
 - **Ex.** Em um caso de internação, 80% das vezes existe necessidade de prescrever soro fisiológico, e que em 20% das vezes, também existe necessidade de prescrever um medicamento para febre.

- **Previsão**

- Permite prever dados futuros partindo de uma série de valores conhecidos.
 - Ex. Dado uma série temporal de batimentos cardíacos, se os valores de pulso continuarem estáveis, quais serão os valores de batimento nos próximos 5 minutos?

Ciência de dados



Machine Learning

Computação cognitiva

- **Conceito**

- A computação cognitiva estende a computação básica, permitindo a máquina inferir e raciocinar, dado um objetivo;
- Capaz de lidar com problemas humanos de forma natural (**agentes de suporte por voz**);
- Através da análise do contexto, obter a melhor resposta ao invés da resposta correta (**sistemas especialistas de busca**).

Computação cognitiva - sistemas devem ser:

- **Adaptativos**

- Um sistema deve ter a capacidade de se **adaptar constantemente** junto aos dados em que é exposto, através da análise em tempo real dos dados.

Computação cognitiva - sistemas devem ser:

- **Adaptativos**

- Um sistema deve ter a capacidade de se **adaptar constantemente** junto aos dados em que é exposto, através da análise em tempo real dos dados.

- **Interativos**

- Um sistema deve se comunicar naturalmente com os usuários, facilitando a interação entre a **máquina** e o **humano**.

Computação cognitiva - sistemas devem ser:

- **Adaptativos**

- Um sistema deve ter a capacidade de se **adaptar constantemente** junto aos dados em que é exposto, através da análise em tempo real dos dados.

- **Interativos**

- Um sistema deve se comunicar naturalmente com os usuários, facilitando a interação entre a **máquina** e o **humano**.

- **Iterativo**

- Um sistema deve ter a capacidade de resolver problemas como ambiguidade ou falta de informação, realizando um conjunto de perguntas para o usuário, assim aumentando o domínio sobre o problema em que se encontra.

Computação cognitiva - sistemas devem ser:

- **Adaptativos**

- Um sistema deve ter a capacidade de se **adaptar constantemente** junto aos dados em que é exposto, através da análise em tempo real dos dados.

- **Interativos**

- Um sistema deve se comunicar naturalmente com os usuários, facilitando a interação entre a **máquina** e o **humano**.

- **Iterativo**

- Um sistema deve ter a capacidade de resolver problemas como ambiguidade ou falta de informação, realizando um conjunto de perguntas para o usuário, assim aumentando o domínio sobre o problema em que se encontra.

- **Contextual**

- Um sistema deve ter a capacidade de utilizar-se do contexto atual do diálogo para ajudar na identificação da melhor resposta para o problema proposto.

Computação cognitiva

- **Como aprendemos**

- Aprendemos de forma muito similar a probabilidade bayesiana.
 - Partimos de uma hipótese;
 - Estabelecemos nossos intervalos de confiança junto ao número de evidências que sustentam nossa hipótese.
- Uma criança, enquanto brinca e explora as coisas que, para ela são novidade, cria hipóteses de como o mundo a sua volta se comporta;
- Realiza testes e verifica se suas hipóteses são corretas;
- Ideias são formuladas e verificadas a cada novo questionamento, brincadeira ou descoberta;
- Desta forma, constrói seus **modelos mentais** de como o mundo funciona;
- Da mesma maneira, fazemos máquinas aprenderem.

Tipos de aprendizado

- **Supervisionado**

- Treinamos o computador com exemplos de dados de entrada e sua saída esperada.
 - **Ex.** Treinamos um modelo para prever o custo de determinado paciente dado seu perfil.
Para isso, treinamos o modelo com nossos dados históricos.

Tipos de aprendizado

- **Supervisionado**

- Treinamos o computador com exemplos de dados de entrada e sua saída esperada.
 - **Ex.** Treinamos um modelo para prever o custo de determinado paciente dado seu perfil. Para isso, treinamos o modelo com nossos dados históricos.

- **Não supervisionado**

- Treinamos o computador sem fornecer nenhum tipo de informação de saída. Esperamos que o computador encontre padrões nos dados fornecidos
 - **Ex.** Esperamos que um modelo encontre dois clusters de pixels dominantes em uma imagem de ressonância magnética.

Tipos de aprendizado

- **Supervisionado**

- Treinamos o computador com exemplos de dados de entrada e sua saída esperada.
 - **Ex.** Treinamos um modelo para prever o custo de determinado paciente dado seu perfil. Para isso, treinamos o modelo com nossos dados históricos.

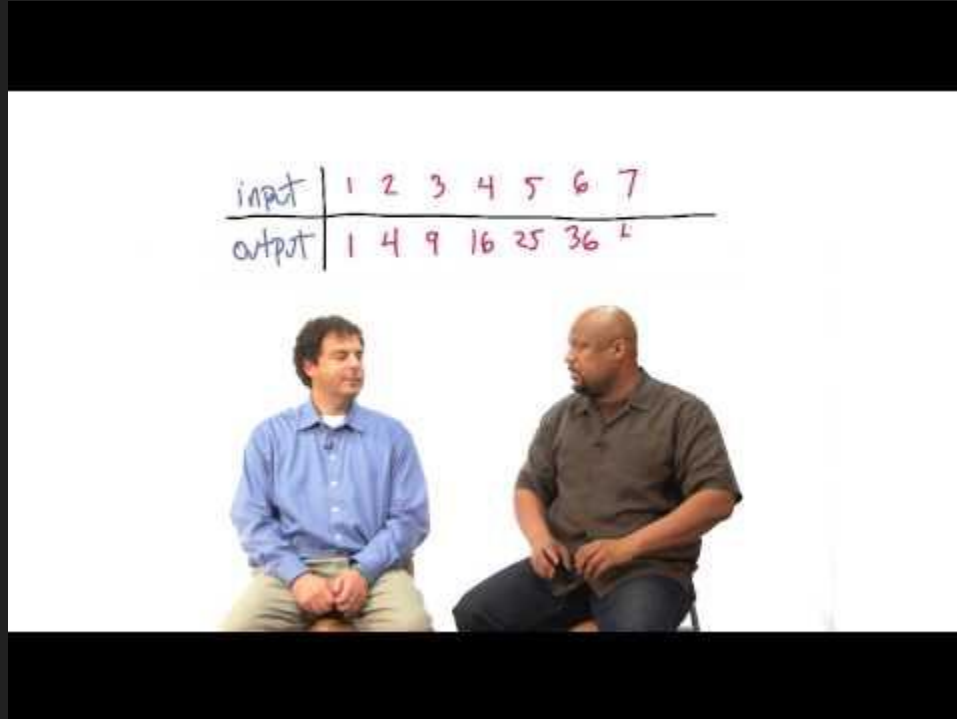
- **Não supervisionado**

- Treinamos o computador sem fornecer nenhum tipo de informação de saída. Esperamos que o computador encontre padrões nos dados fornecidos
 - **Ex.** Esperamos que um modelo encontre dois clusters de pixels dominantes em uma imagem de ressonância magnética.

- **Por reforço**

- Modelo é exposto ao ambiente, seja ele real ou simulado. Espera-se que o modelo execute alguma função e fornecemos um resultado ao modelo no final de cada ação.
 - **Ex.** Esperamos que o modelo aprenda a melhor forma de distribuir demandas de pacientes no pronto socorro aos médicos seguindo uma regra de negócio.

Tipos de aprendizado - Supervisionado



Tipos de aprendizado - Não supervisionado



Tipos de aprendizado - Por reforço

