

Tópicos

Avançados II

Aula 8

FRAMEWORKS & TOOLS TO MACHINE LEARNING

AUGUSTO UCHÔA



Trilha de hoje????

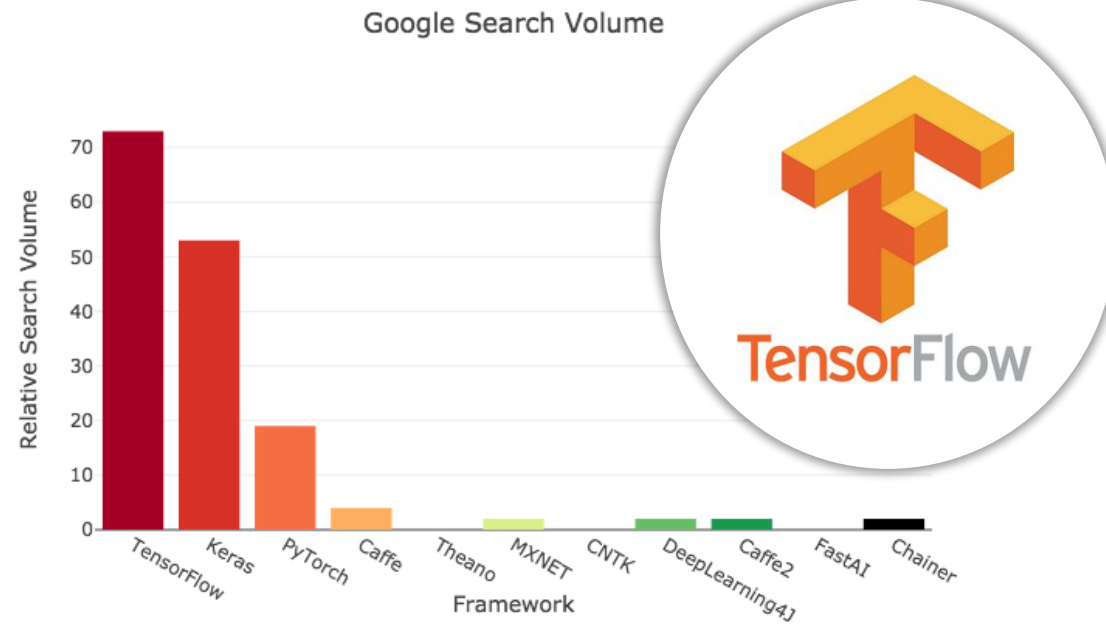
Conhecer as principais Frameworks, Tools e “bibliotecas” para Machine Learning

Escolher a partir de características pessoais e objetivos em modelagem neural qual ferramenta usar

Perceber que existem alternativas mesmo para quem nunca escreveu uma linha de código

Experimentar o Bing CHAT e o Matlab para o pré-processamento de um dataset com fins de modelagem neural

O TensorFlow é uma plataforma de **código aberto** para aprendizado de máquina e modelagem neural. Desenvolvido pela equipe do **Google Brain**, é uma das ferramentas mais utilizadas em todo o mundo.



Desvantagens

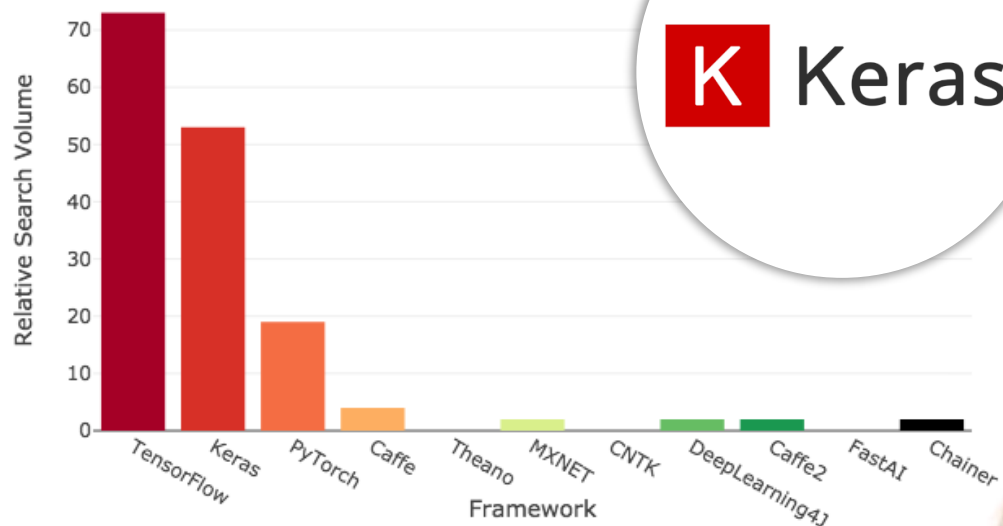
- Limitações na customização de modelos
- Problemas de performance com grandes conjuntos de dados
- Necessita de conhecimento prévio em aprendizado de máquina



Vantagens

- Facilidade de uso
- Grande comunidade de usuários
- Compatibilidade com várias linguagens de programação

Google Search Volume



 **Keras**

O Keras é uma API de alto nível para construção e treinamento de modelos de aprendizado de máquina em Python. Ele é um wrapper para outras bibliotecas de computação numérica, como o TensorFlow, Pytorch e o Theano.

Vantagens

- Facilidade de uso e de customização de modelos
- Sintaxe simples e elegante
- Compatibilidade com várias bibliotecas de computação numérica

Desvantagens

- Limitações em modelos mais complexos
- Nem sempre é a melhor escolha para tarefas específicas
- Dependência de outras bibliotecas

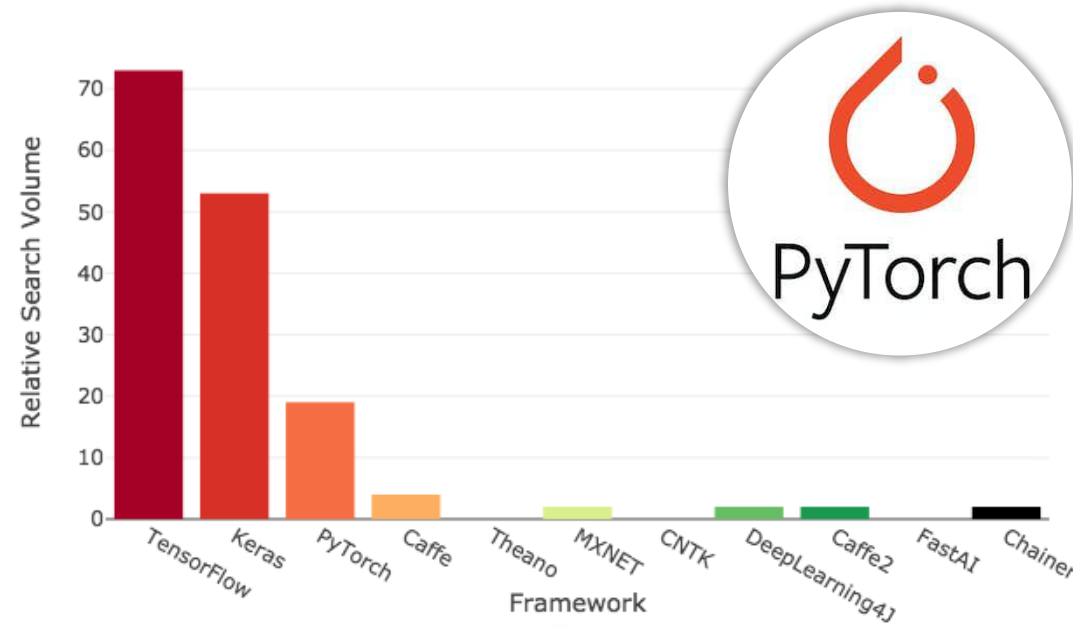


O PyTorch é outra **plataforma de código aberto** para aprendizado de máquina, desenvolvida pela equipe do **Facebook AI**. Ela utiliza um sistema de gráficos dinâmicos para acelerar o processo de treinamento de modelos.

Vantagens

- Facilidade de uso e de customização de modelos
- Performance superior ao TensorFlow em algumas situações
- Comunidade de usuários em franca expansão

Google Search Volume



Desvantagens

- Menor compatibilidade com outras linguagens de programação
- Necessita de recursos computacionais mais poderosos para treinamento de modelos complexos
- Curva de aprendizado mais complexa do que o TensorFlow

O Scikit Learn é uma biblioteca Python voltada para aprendizado de máquina. Ele conta com ferramentas simples e eficientes para modelagem de dados e análise estatística.



Vantagens

- Simplicidade de uso
- Ampla documentação disponível na internet
- Compatibilidade com outros módulos Python

Desvantagens

- Menos recursos e funcionalidades avançadas do que algumas outras ferramentas
- Limitações na eficiência no treino de modelos mais complexos
- Incompatibilidade com outras linguagens de programação



<https://youtu.be/eqaskloXzgY?list=RDCMUCLEQMN2aaUD9VfBVPxG0IOg>

Canal Caio Dallaqua no Youtube

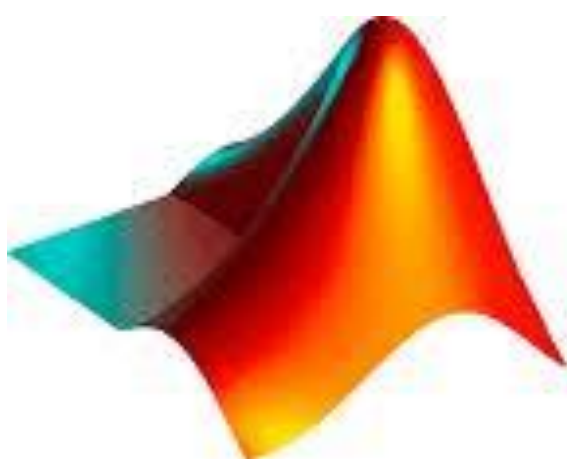
VOCÊ JÁ USA O PHYTON?

SE VOCÊ NÃO PROGRAMA NEM OS CANAIS DA TV, A DICA É OUTRA:



Canal OPENCADD
no Youtube

<https://youtu.be/i2gX28Y7PeU>



MATLAB

O MATLAB é uma plataforma numérica de alto nível que permite realizar cálculos complexos e gerar visualizações interativas. Ele conta com uma toolbox específica para aprendizado de máquina, que vem se popularizando nos últimos anos.



Vantagens

- Facilidade de uso e de customização de modelos
- Sintaxe elegante e fácil de entender
- Bom desempenho em tarefas específicas

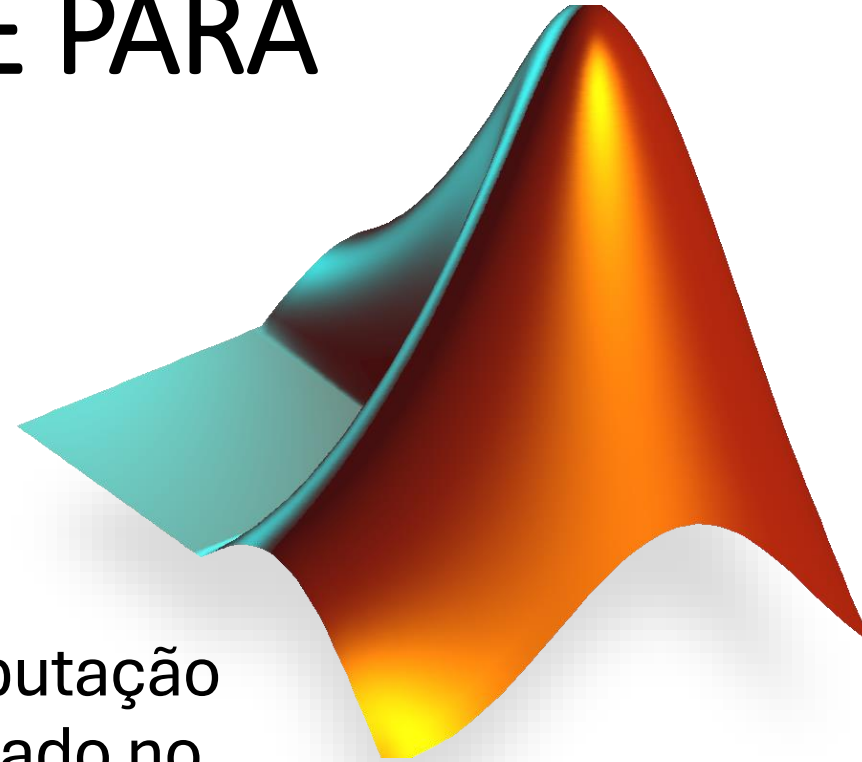
Desvantagens

• **Custo**

- Recursos computacionais poderosos necessários para treinamento de modelos complexos
- Nem sempre é a melhor escolha para tarefas mais amplas

O MATLAB É FREE PARA ALUNOS DA UFC

MATLAB é o software de computação numérico-científica mais utilizado no mundo pelas maiores e melhores universidades. O software possui linguagem simples e direta para realização de cálculos matemáticos, possui alta performance computacional e possui uma ampla biblioteca de funções matemáticas pré-definidas, incluindo de neural networks e Deep Learning.



INFORMAÇÃO IMPORTANTE PARA ALUNOS ACESSO AO MATLAB



UFC oferece, aos seus alunos, licenças MATLAB e Simulink para download e uso em seus computadores pessoais.

Esses softwares são necessários para a realização de trabalhos acadêmicos de várias disciplinas.

Siga as instruções e faça seu download.

Instruções de acesso:

1. Acesse [UFC Portal MATLAB](#) para fazer o download do software
2. Clique em "Sign in to get started" na seção Get MATLAB.
3. Será solicitado que você crie uma *MathWorks Account*. Uma vez criada a conta, seu perfil será associado à Campus-Wide License da UFC e você poderá:
 - Baixar e ativar os softwares em seu computador pessoal
 - Utilizar o MATLAB Online diretamente do browser da internet

[Acesse o UFC Portal MATLAB](#)

**MATLAB®
& SIMULINK®**



Informações adicionais:

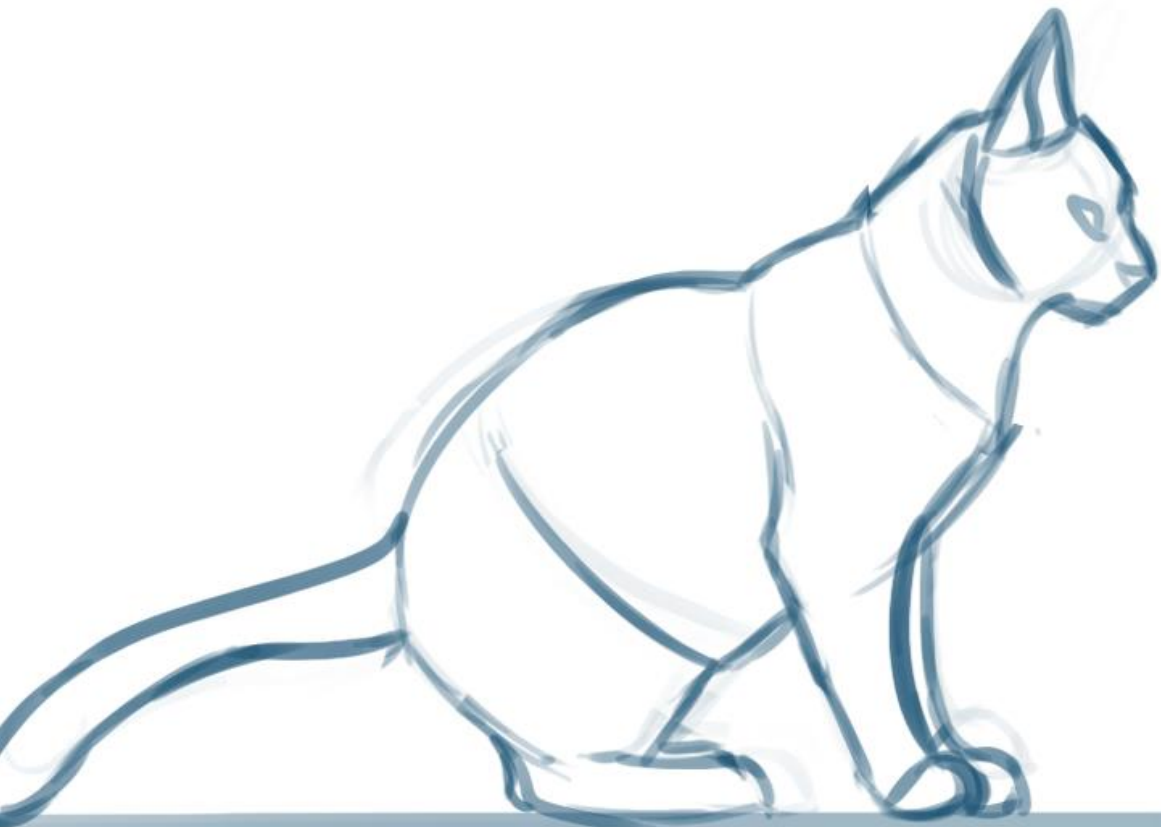
Caso você não saiba usar o MATLAB, acesse o [MATLAB Onramp](#), um tutorial com duração de 2 horas, que pode ser encontrado em [UFC Portal MATLAB](#) na seção Learn to use MATLAB and Simulink.

Se você tiver qualquer problema durante a instalação do MATLAB, acesse [PUCPR Portal MATLAB](#) e clique em "Need Installation Help" para solicitar suporte técnico.

O PULO DO



Trabalha com LLMs - Large Language Models



Não programa nem os canais da TV



ARTIFICIAL
INTELLIGENCE



**PARA
AJUDAR
NO
MATLAB:**

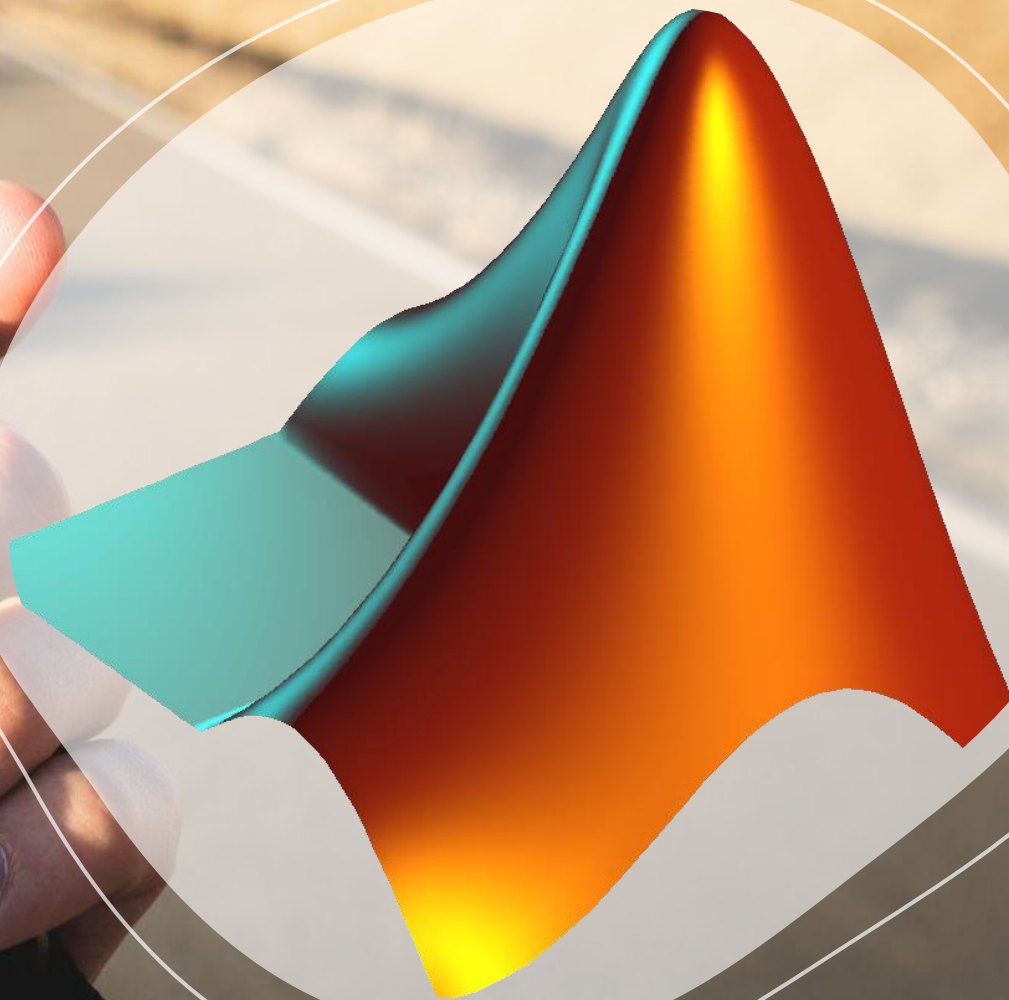
M é uma linguagem de programação de alto nível que é projetada para expressar matemática de matriz. O MATLAB fornece milhares de funções embutidas para cálculos matemáticos, científicos e de engenharia comuns. Mas você não precisa saber programar em M, só precisa saber “promptar”

Partiu
Garotinha!

HOJE NÃO
TEM
ATIVIDADE!
MAS TEM
PRÁTICA



**PARTIU
MATLAB!**



ENTENDA O FENÔMENO:

1. O fenômeno modelado corresponde a um tema investigado em Infraestrutura de Transportes, linha de gerência de pavimentos, Projeto de Gerência de Pavimentos Aeroportuários. Deseja-se atribuir as classes de: segura ou não segura à PPD no Brasil em função de parâmetros de aderência pneu/pavimento.
2. Como este é o primeiro modelo neural da vida de vocês, decidi simplificar as coisas e não aplicaremos nenhuma técnica de regularização (dropout) e nem a técnica K-fold (Validação cruzada). Simplificadamente dividiremos o Dataset disponível em apenas 2 conjuntos, a saber:
 - a) Treinamento com 70% dos dados) e;
 - b) Teste com 30% dos dados.
3. Criou-se no Excel, uma planilha onde foram inseridos os dados do fenômeno a ser modelado (complexo, multivariado e não linear (representatividade e equilíbrio das classes é outra história);
4. O Dataset é composto de 6 variáveis independentes e 1 dependente;
5. Cada linha da planilha representa um vetor onde as 6 primeiras colunas são as variáveis independentes (entradas) e a 7ª coluna a variável dependente (saída);
6. Após popular o banco com dados obtidos em observações ou ensaios ou ainda dados secundários (coletados por outrem)
7. Por exemplo, salvou-se esse arquivo em formato de texto separado por tabulações (Dataset aula.txt).



Entradas

Saídas

ID	variável entrada 1	variável entrada 2	variável entrada 3	variável entrada 4	variável entrada 5	variável entrada 6	variável saída
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
.							
.							
.							
1238							

Padrões

[Baixe o Dataset para prática no Bing CHAT e Matlab-Clique aqui!](#)

AO INSTALAR O MATLAB!

Não esqueça de selecionar **“Select All”** para instalação e todos os toolboxes, para trabalhar com Neural Networks, machine Learning e Deep Learning precisa de diversos toolboxes, marque tudo para prevenir pois nunca se sabe o que vai precisar.

The screenshot shows the MathWorks Product Installer for R2023b. The installer is in the 'PRODUCTS' step, where users can select products to install. The 'Select All' option is circled in red, indicating that all recommended products and toolboxes should be selected for installation.

Product	Selected
Select All	<input checked="" type="checkbox"/>
MATLAB	<input checked="" type="checkbox"/>
Simulink	<input checked="" type="checkbox"/>
5G Toolbox	<input checked="" type="checkbox"/>
Aerospace Blockset	<input checked="" type="checkbox"/>
Aerospace Toolbox	<input checked="" type="checkbox"/>
Antenna Toolbox	<input checked="" type="checkbox"/>
Audio Toolbox	<input checked="" type="checkbox"/>
Automated Driving Toolbox	<input checked="" type="checkbox"/>
AUTOSAR Blockset	<input checked="" type="checkbox"/>
Bioinformatics Toolbox	<input checked="" type="checkbox"/>
Bluetooth Toolbox	<input checked="" type="checkbox"/>

NO EXCEL

1. Criei uma pasta na sua área de trabalho chamada “dados excel matlab”
2. Inseriria nesta pasta o arquivo excel que contem seus dados, chamado de “dadosexcel”;
3. Entenda o que tem nesse arquivo do excel:
 - a) uma planilha contendo as variáveis independentes (entradas) nas colunas B, C, D, E, F e G. a variável dependente (saída) na coluna H. São 1238 linhas (padrões) por 7 colunas de dados efetivos;

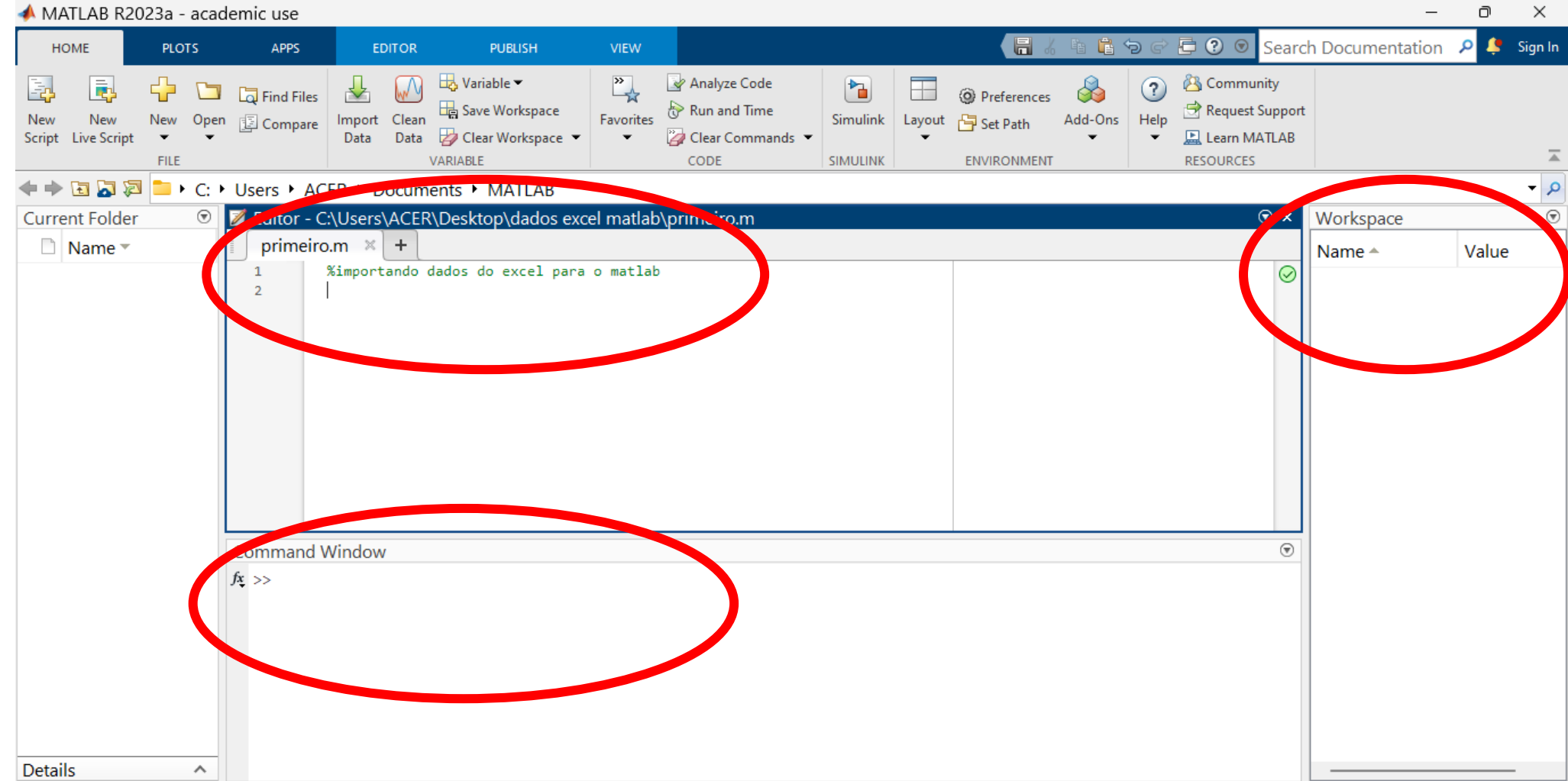
	A	B	C	D	E	F	G	H
1	ID	CA	PM	RAB	TP	LP	RV	CSO
2	0	0,64	0,62	1,00	1,00	100,00	0,00	0,00
3	1	0,76	0,69	1,00	1,00	100,00	0,00	0,00
4	2	0,64	0,62	1,00	1,00	100,00	0,00	0,00
5	3	0,62	0,79	0,00	1,00	100,00	0,00	0,00
6	4	0,78	0,88	1,00	1,00	100,00	0,00	1,00
7	5	0,67	0,63	1,00	1,00	100,00	0,00	0,00
8	6	0,70	0,67	1,00	1,00	100,00	0,00	0,00
9	7	0,79	0,71	1,00	1,00	100,00	0,00	0,00
10	8	0,63	0,82	1,00	1,00	100,00	0,00	0,00
11	9	0,65	0,78	1,00	1,00	100,00	0,00	0,00
12	10	0,65	1,06	0,00	1,00	100,00	0,00	0,00
13	11	0,86	0,85	1,00	1,00	100,00	0,00	1,00
14	12	0,78	0,78	1,00	1,00	100,00	0,00	0,00
15	13	0,59	0,69	1,00	1,00	100,00	0,00	0,00
16	14	0,68	0,77	0,00	1,00	100,00	0,00	0,00
17	15	0,54	0,92	0,00	1,00	100,00	0,00	0,00
18	16	0,56	1,18	1,00	1,00	100,00	0,00	0,00

1224	1222	0,57	1,27	1,00	3,00	1000,00	1,00	0,00
1225	1223	0,61	1,37	1,00	3,00	1000,00	1,00	0,00
1226	1224	0,61	1,38	1,00	3,00	1000,00	1,00	0,00
1227	1225	0,63	1,22	1,00	3,00	1000,00	1,00	0,00
1228	1226	0,61	1,22	1,00	3,00	1000,00	1,00	0,00
1229	1227	0,70	1,16	1,00	3,00	1000,00	1,00	0,00
1230	1228	0,67	1,32	0,00	3,00	1000,00	1,00	1,00
1231	1229	0,57	1,40	1,00	3,00	1000,00	1,00	0,00
1232	1230	0,78	1,40	1,00	3,00	1000,00	1,00	1,00
1233	1231	0,63	1,45	1,00	3,00	1000,00	1,00	0,00
1234	1232	0,69	1,57	1,00	3,00	1000,00	1,00	1,00
1235	1233	0,63	1,39	1,00	3,00	1000,00	1,00	0,00
1236	1234	0,67	1,54	1,00	3,00	1000,00	1,00	1,00
1237	1235	0,67	1,34	1,00	3,00	1000,00	1,00	1,00
1238	1236	0,58	1,58	1,00	3,00	1000,00	1,00	0,00
1239	1237	0,62	1,65	1,00	3,00	1000,00	1,00	0,00

MATLAB, muito prazer!

•**Script:** É um arquivo que contém uma sequência de comandos MATLAB. Você pode executar um script digitando seu nome na janela de comando. Os scripts são úteis para automatizar tarefas repetitivas e para executar uma série de comandos em ordem.

•**Workspace:** É uma área de memória onde o MATLAB armazena variáveis e seus valores. Você pode visualizar o conteúdo do workspace digitando o comando whos na janela de comando. O workspace é útil para armazenar dados temporários ou permanentes.

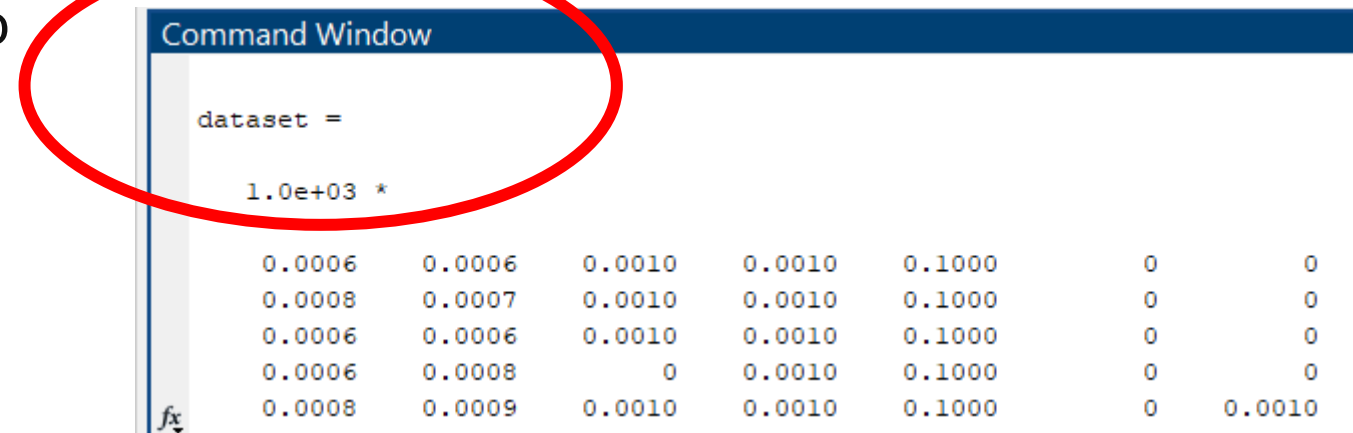
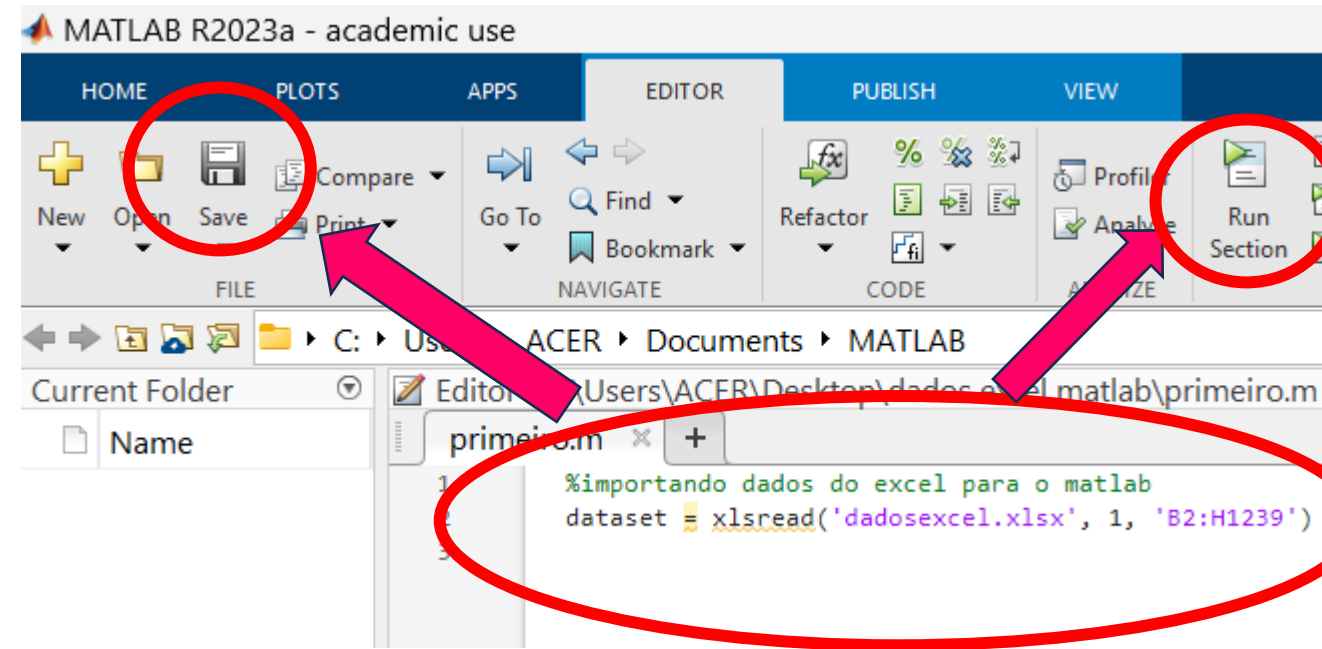


•**Command Window:** É a janela onde você digita comandos MATLAB e recebe saídas. O Command Window é útil para testar comandos individuais, depurar código e executar scripts.

No Dataset = xlsread('dadosexcel.xlsx',1,'B2-H1239')

MATLAB

- Escreva seu primeiro script:
Na janela de script escreva:
- % importando dados do excel no matlab, e salve este script para não perder, com o nome de “primeiro”;
- É IMPORTANTE que o script seja salvo na mesma pasta que o arquivo do excel que será importado, ou seja: na pasta que está na área de trabalho chamada de “dados excel matlab”



SUA SITUAÇÃO

Possuo um arquivo, chamado Dataset aula.txt, em formato de texto separado por tabulações, contendo 1238 linhas e 7 colunas

SEU OBJETIVO

Preciso carregar esse arquivo no matlab, depois preciso reduzir a amplitude desses dados para o intervalo entre 0 e 1, além disso, preciso randomizar a ordem das linhas desse arquivo e, finalmente separar esses dados em 4 arquivos, quais sejam: o primeiro arquivo deve conter 70% dos dados e nele devem estar contidas as 6 primeiras colunas do arquivo e deve ser nomeado de entradastreino, o segundo arquivo deve obedecer à mesma estrutura do anterior, mas deve conter apenas os dados da sétima coluna e deve ser nomeado de saidastreino. O quarto conjunto deve conter os 30% restantes após a randomização e, deve conter os dados contidos da primeira à sexta coluna e deve ser nomeado de entrasteste e finalmente o quarto conjunto deve obedecer à mesma estrutura do terceiro, mas nele devem estar contidos apenas os dados relativos à sétima coluna e deve ser nomeado de saidasteste.

ATUE/ SEJA

Seja um especialista em matlab e machine learning com 10 anos de experiência e didática para capacitação de iniciantes

TAREFA

Crie uma lista, estilo passo a passo, incluindo os códigos na linguagem m, para uso no matlab, detalhando cada passo e sua utilidade.

ESTRUTURE SEU PROMPT



após ter carregado um arquivo excel no matlab com o comando: `dataset = xlsread('dadosexcel.xlsx', 1, 'B2:H1239')`, preciso realizar as seguintes tarefas e preciso de código no matlab para: criar uma ordem randômica das 1238 linhas, normalizar todos os valores usando a função mini-max entre 0 e 1, separar em 2 conjuntos de dados mutuamente exclusivos, o primeiro deve conter 70% dos dados e deve ser nomeado de treinamento, o segundo deve conter o restante dos dados e ser nomeado de teste. Agora cada um desses conjuntos devem ser divididos em dois subconjuntos, a saber: `entradastreinamento`, nele devem estar contidas os dados da coluna B até a coluna G, e `saidatreinamento`, deste deve conter os dados da coluna H. De forma análoga o conjunto nomeado de teste deve ser dividido em dois conjuntos, a saber: `entradasteste`, nele devem estar contidas os dados da coluna B até a coluna G, e `saidateste`, deste deve conter os dados da coluna H. Me forneça o código em m contendo comentários explicativos de cada comando.

**No
Bing
CHAT**

“após ter carregado um arquivo excel no matlab com o comando: `dataset = xlsread('dadosexcel.xlsx', 1, 'B2:H1239')`, preciso realizar as seguintes tarefas e preciso de código no matlab para: criar uma ordem randômica das 1238 linhas, normalizar todos os valores usando a função mini-max entre 0 e 1, separar em 2 conjuntos de dados mutuamente exclusivos, o primeiro deve conter 70% dos dados e deve ser nomeado de treinamento, o segundo deve conter o restante dos dados e ser nomeado de teste. Agora cada um desses conjuntos devem ser divididos em dois subconjuntos, a saber: `entradastreinamento`, nele devem estar contidas os dados da coluna B até a coluna G, e `saidatreinamento`, deste deve conter os dados da coluna H. De forma análoga o conjunto nomeado de teste deve ser dividido em dois conjuntos, a saber: `entradasteste`, nele devem estar contidas os dados da coluna B até a coluna G, e `saidateste`, deste deve conter os dados da coluna H. Me forneça o código em m contendo comentários explicativos de cada comando.”



No Bing CHAT

Olá! Para realizar as tarefas que você mencionou, você pode usar o seguinte código em MATLAB:

```
% Carregando os dados do arquivo Excel
dataset = xlsread('dadosexcel.xlsx', 1, 'B2:H1239');

% Criando uma ordem randômica das linhas
randomOrder = randperm(size(dataset, 1));
dataset = dataset(randomOrder, :);

% Normalizando todos os valores usando a função mini-max entre 0 e 1
datasetNormalized = (dataset - min(dataset)) ./ (max(dataset) - min(dataset));

% Separando em dois conjuntos de dados mutuamente exclusivos
trainingSetSize = round(size(datasetNormalized, 1) * 0.7);
trainingSet = datasetNormalized(1:trainingSetSize, :);
testSet = datasetNormalized(trainingSetSize+1:end, :);

% Dividindo cada conjunto em dois subconjuntos
entradasTreinamento = trainingSet(:, 1:6);
saidasTreinamento = trainingSet(:, 7);
entradasTeste = testSet(:, 1:6);
saidasTeste = testSet(:, 7);
```

Editor - C:\Users\ACER\Desktop\dados excel matlab\primeiro.m

```
3
4 % Criando uma ordem randômica das linhas
5 randomOrder = randperm(size(dataset, 1));
6 dataset = dataset(randomOrder, :);
7
8 % Normalizando todos os valores usando a função mini-max entre 0 e 1
9 datasetNormalized = (dataset - min(dataset)) ./ (max(dataset) - min(dataset));
10
11 % Separando em dois conjuntos de dados mutuamente exclusivos
12 trainingSetSize = round(size(datasetNormalized, 1) * 0.7);
13 trainingSet = datasetNormalized(1:trainingSetSize, :);
14 testSet = datasetNormalized(trainingSetSize+1:end, :);
15
16 % Dividindo cada conjunto em dois subconjuntos
17 entradasTreinamento = trainingSet(:, 1:6);
18 saidasTreinamento = trainingSet(:, 7);
19 entradasTeste = testSet(:, 1:6);
20 saidasTeste = testSet(:, 7);
21
```

Workspace

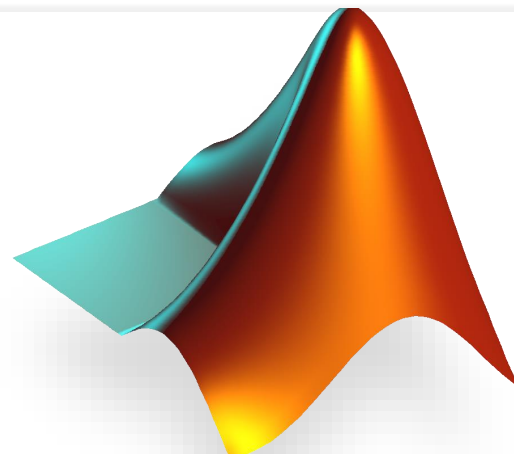
Name ^	Value
dataset	1238x7 double
datasetNorm...	1238x7 double
entradasTeste	371x6 double
entradasTrein...	867x6 double
randomOrder	1x1238 double
saidasTeste	371x1 double
saidasTreina...	867x1 double
testSet	371x7 double
trainingSet	867x7 double
trainingSetSize	867

Command Window

```
0.0007    0.0015    0.0010    0.0030    1.0000    0.0010    0.0010
0.0007    0.0013    0.0010    0.0030    1.0000    0.0010    0.0010
0.0006    0.0016    0.0010    0.0030    1.0000    0.0010         0
0.0006    0.0016    0.0010    0.0030    1.0000    0.0010         0
```

fx >>

No
MATLAB



PULGA ATRÁS DA ORELHA?

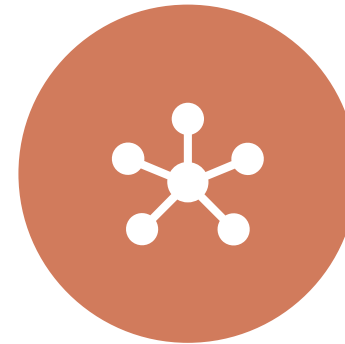
1. Deus, será que consigo mesmo trabalhar no Matlab?
2. Deu certo meu pré-processamento?
3. ?????



PONTOS CHAVE



Principais Frameworks and tools para Machine Learning



Recomendações para que já programa em Python



Recomendações para quem não programa nem o controle remote da TV



Prática no Matlab/Bing CHAT, para pré-processamento de dados com fins de modelagem neural

A SEGUIR, CENAS DO
PRÓXIMO
CAPÍTULO

**PRÁTICA DE
MODELAGEM
NEURAL NO
MATLAB**

