

Actividad #14

Table of Contents

Fecha:	1
Objetivos:	1
Nombre:	1
Repository:	2
Libraries:	2
Paso 0: Descartar cualquier cambio realizado en el repositorio clonado	2
Paso 1: Limpiar variables y linea de comandos	3
Paso 2.- Configuración de carpeta ./src para librerías	3
Paso 3- Configuración de carpeta de ./data para datasets	3
Paso 4- Buscar los nombres y Cargar los datos de todos los archivos dentro de la carpeta ./data	3
Paso 5- Extraer nombres de variables y crear datetime	4
Paso 6- Graficar todas las variables	4
Paso 7- Definir la variable de salida y las variables de entrada del sistema	6
Paso 8- Seleccionar Variables o características	7
Paso 9- Seleccionar el algoritmo de ML con un menos error de prediccion empleando el toolbox de Matlab Regression Learner	9
Paso 10- Dividir el dataset en 70% para entrenar y 30% validar	11
Paso 11- Usando el algoritmo de ML se entrena el modelo de regression (costo computacional)	12
Paso 12- Cargar y validar el modelo entrenado	13
paso 13 - Graficar el valor predecido vs el valor real	13
paso 14 - Mejorar el modelo de prediccion	14

Fecha:

```
fecha = datetime('now', 'Format', 'dd-MM-yyyy');  
disp(['Fecha actualizada: ', char(fecha)])
```

Fecha actualizada: 06-07-2024

Objetivos:

- Leer datos desde el repositorio: Open Energy Data Initiative
- https://openei.org/datasets/files/961/pub/COMMERCIAL_LOAD_DATA_E_PLUS_OUTPUT/
- Definir la variable de salida y las variables de entrada del sistema
- Seleccionar Variables o características empleando la matriz de correlacion
- Seleccionar el algoritmo de ML con un menos error de prediccion empleando el toolbox de Matlab Regression Learner
- Dividir el dataset en 70% para entrenar y 30% validar
- Usando el algoritmo de ML se entrena el modelo de regression (costo computacional)
- Cargar y validar el modelo entrenado
- Graficar el valor predecido vs el valor real

Nombre:

- sunombre

Repository:

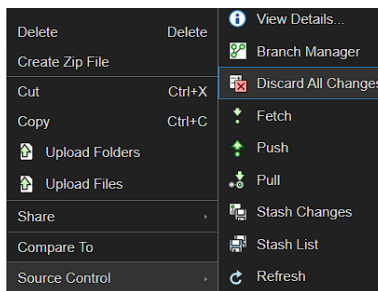
- <https://github.com/vasanza/SSE>

Librarys:

- https://github.com/vasanza/Matlab_Code
- <https://es.mathworks.com/help/matlab/ref/lis.htm>
- <https://es.mathworks.com/help/matlab/ref/matlab.git.gitrepository.discardchanges.html#d126e406558>

Paso 0: Descartar cualquier cambio realizado en el repositorio clonado

```
%Version Online, Opcion 1:  
% Source Control -> Discard all changes  
% Source Control -> git pull
```



```
%Version Online, Opcion 2:  
% repo = gitrepo;  
% discardChanges(repo,repo.ModifiedFiles);  
% Source Control -> git pull
```

```
Command Window  
>> repo = gitrepo;  
discardChanges(repo,repo.ModifiedFiles);  
>>
```

```
% Version para PC, en el Bash del Git:  
% git status  
% git reset --hard  
% Git pull
```

```

MINGW64 ~/Users/ALCORT/SSSE/2024
ALCORT@DESKTOP-KHONPRN MINGW64 ~/SSSE/2024 (main)
$ git status
On branch main
Your branch is behind 'origin/main' by 5 commits, and can be fast-forwarded.
(use "git pull" to update your local branch)

Changes not staged for commit:
  (use "git add <files> ..." to update what will be committed)
  (use "git restore <files> ..." to discard changes in working directory)
        modified:   ACTIVIDAD10/main.mlx
        modified:   ACTIVIDAD09/main - Copy.mlx
        modified:   ACTIVIDAD09/main.mlx

no changes added to commit (use "git add" and/or "git commit -a")

ALCORT@DESKTOP-KHONPRN MINGW64 ~/SSSE/2024 (main)
$ git reset --hard
HEAD is now at 5ae2883 Add files via upload

ALCORT@DESKTOP-KHONPRN MINGW64 ~/SSSE/2024 (main)
$ git pull
Updating 5ae2883..c69ecb8
Fast-forward
 2024/ACTIVIDAD10/data/Cliente1/2024-06-25.csv | 2814 ++++++
 2024/ACTIVIDAD10/data/Cliente1/2024-06-26.csv | 2820 ++++++
 2024/ACTIVIDAD10/data/Cliente1/2024-06-27.csv | 2820 ++++++
 2024/ACTIVIDAD10/data/Cliente1/2024-06-28.csv | 2820 ++++++
 2024/ACTIVIDAD10/data/Cliente2/2024-06-2.csv | 2820 ++++++
 2024/ACTIVIDAD10/data/Cliente2/2024-06-21.csv | 2814 ++++++
 2024/ACTIVIDAD10/data/Cliente2/2024-06-22.csv | 2820 ++++++
 2024/ACTIVIDAD10/data/Cliente2/2024-06-23.csv | 2820 ++++++
 2024/ACTIVIDAD10/data/Cliente3/2023-06-21.csv | 2814 ++++++
 2024/ACTIVIDAD10/data/Cliente3/2023-06-22.csv | 2820 ++++++
 2024/ACTIVIDAD10/data/Cliente3/2023-06-23.csv | 2820 ++++++
 .../data/Cliente3/Copy_of_2023-06-23.csv | 2820 ++++++
 .../data/Copy_2_of_Cliente3/2023-06-21.csv | 2814 ++++++
 .../data/Copy_2_of_Cliente3/2023-06-22.csv | 2820 ++++++
 .../data/Copy_2_of_Cliente3/2023-06-23.csv | 2820 ++++++
 2024/ACTIVIDAD10/main.mlx | Bin 255418 -> 52128 bytes
 2024/ACTIVIDAD10/src/Plot.m | 31 +
 18 files changed, 22579 insertions(+), 22548 deletions(-)
 create mode 100644 2024/ACTIVIDAD10/src/Plot.m

ALCORT@DESKTOP-KHONPRN MINGW64 ~/SSSE/2024 (main)
$

```

Paso 1: Limpiar variables y linea de comandos

```

clear % Para borrar el workspace y liberar memoria RAM
clc % Limpiar el command window

```

Paso 2.- Configuración de carpeta ./src para librerias

```

%nombre de la carpeta donde estan los codigos
addpath(genpath('./src'));

```

Paso 3- Configuración de carpeta de ./data para datasets

```

%Nombre de la carpeta donde estan los archivos csv
datapath=fullfile('./data/');

```

Paso 4- Buscar los nombres y Cargar los datos de todos los archivos dentro de la carpeta ./data

```

%Leer un archivo csv y lo carga como una tabla
filename = FindCSV(datapath);
% Dataset es una tabla donde cada columna es una variable con su
% respectivo nombre
Dataset=fLoadTableCSV(filename,datapath)

```

Warning: Column headers from the file were modified to make them valid MATLAB identifiers before creating variable names for the table. The original column headers are saved in the VariableDescriptions property. Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable names.

Dataset = 8760x11 table

...

	Date_Time	Electricity_Facility_kW_Hourly_	Fans_Electricity_kW_Hourly_
1	01/01/2024 ...	22.0360	3.5862
2	01/01/2024 ...	14.6498	0

	Date_Time	Electricity_Facility_kW__Hourly__	Fans_Electricity_kW__Hourly__
3	01/01/2024 ...	14.6696	0
4	01/01/2024 ...	14.6778	0
5	01/01/2024 ...	14.8248	0
6	01/01/2024 ...	22.1826	3.5862
7	01/01/2024 ...	38.1318	3.5862
8	01/01/2024 ...	45.5973	3.5862
9	01/01/2024 ...	45.6036	3.5862
10	01/01/2024 ...	38.1138	3.5862
11	01/01/2024 ...	42.4352	3.5862
12	01/01/2024 ...	42.1437	3.5862
13	01/01/2024 ...	41.4196	3.5862
14	01/01/2024 ...	33.8970	3.5862

⋮

Paso 5- Extraer nombres de variables y crear datetime

```
% Extraer todos los nombres de variables de la tabla
% Se hace el cast de cell a string
varnames=string(Dataset.Properties.VariableNames);
% Eliminar el primero nombbre de variable
varnames=varnames(2:end)';

% Esto es para eliminar el warning de los legend en el plot
%LegendNames=char(varnames);
%LegendNames=LegendNames(:,1:15);
%LegendNames=[LegendNames char(65*ones([size(varnames,1),1]))];
%LegendNames=string(LegendNames);

% Crear datetime con una frecuencia de muestreo de un dato por hora segun el
% dataset
%Time = Start Time: Step Time: End Time
time = datetime(2004, 1, 1):hours(1):datetime(2004, 12, 31);
% Se elimina el primer valor
time=time(1,2:end)';

% para agregar una nueva variables en la tabla
%Dataset.('Time_Stamp')=time;
```

Paso 6- Graficar todas las variables

```
figure;
% Dataset
```

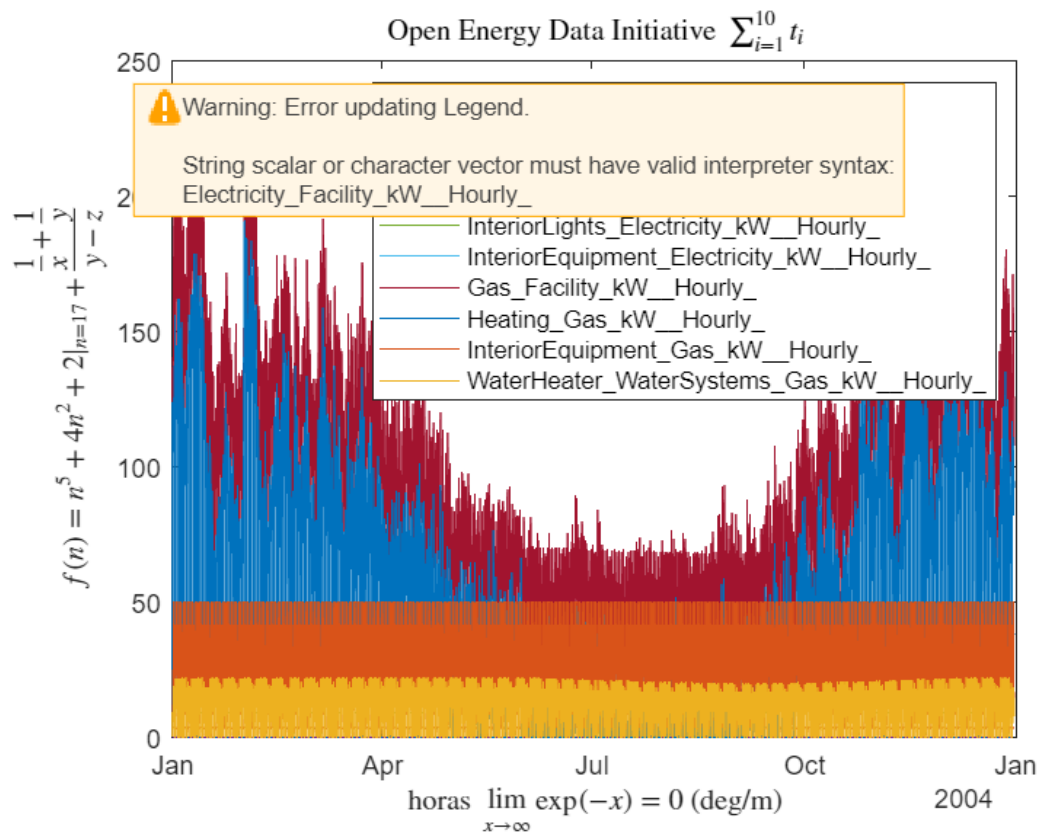
```

Variable1=Dataset.(varnames(1));
plot(time,Variable1)
hold on
Variable2=Dataset.(varnames(2));
plot(time,Variable2)
Variable3=Dataset.(varnames(3));
plot(time,Variable3)
Variable4=Dataset.(varnames(4));
plot(time,Variable4)
Variable5=Dataset.(varnames(5));
plot(time,Variable5)
Variable6=Dataset.(varnames(6));
plot(time,Variable6)
Variable7=Dataset.(varnames(7));
plot(time,Variable7)
Variable8=Dataset.(varnames(8));
plot(time,Variable8)
Variable9=Dataset.(varnames(9));
plot(time,Variable9)
Variable10=Dataset.(varnames(10));
plot(time,Variable10)

hold off
%legend(LegendNames);
legend(varnames);

% tambien se puede usar latex para los lables y titles
% https://en.wikibooks.org/wiki/LaTeX/Mathematics
title('Open Energy Data Initiative $ \sum_{i=1}^{10} t_i $', 'interpreter', 'latex');
xlabel('horas $ \lim\limits_{x \to \infty} \exp(-x) = 0 $ (deg/m)', 'interpreter', 'latex')
ylabel(' $ f(n) = n^5 + 4n^2 + 2 \mid_{n=17} + \frac{\frac{1}{x}+\frac{1}{y}}{y-z} $ ', 'interpreter', 'latex')

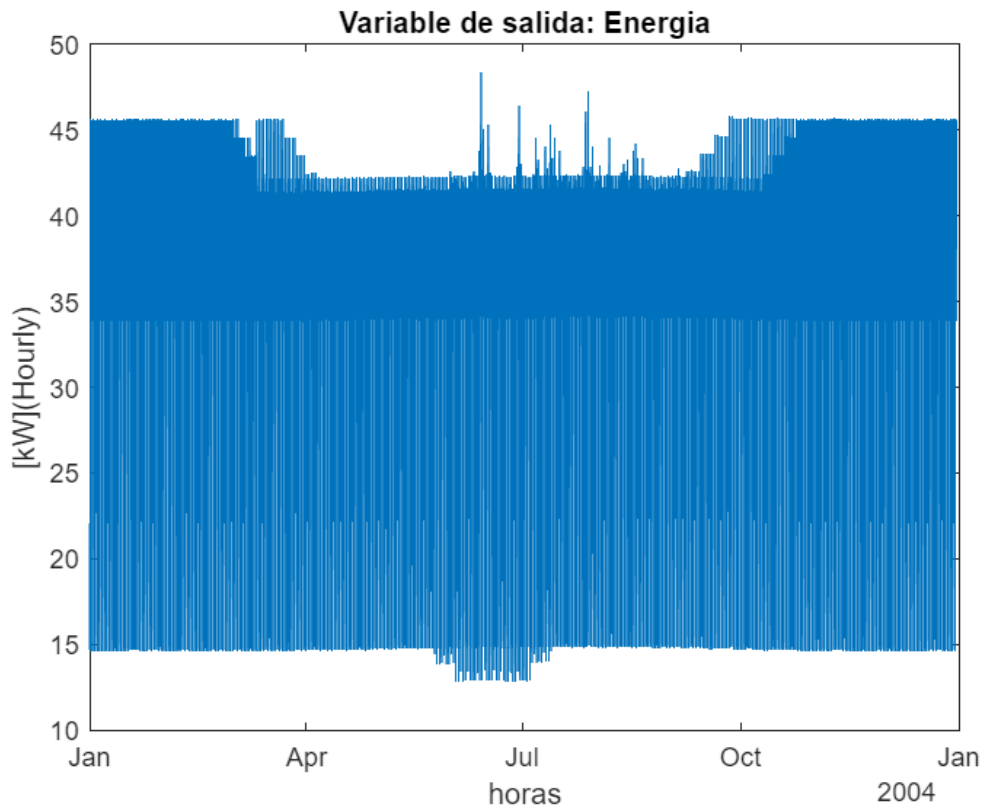
```



```
clear Variable1 Variable2 Variable3 Variable4 Variable5 Variable6 ...  
Variable7 Variable8 Variable9 Variable10
```

Paso 7- Definir la variable de salida y las variables de entrada del sistema

```
figure;  
% Dataset  
output=Dataset.(varnames(1));  
plot(time,output)  
title('Variable de salida: Energia');  
xlabel('horas')  
ylabel('[kW](Hourly)')
```



```
% Estamos usando las variables como características
input=[Dataset.(varnames(2)) Dataset.(varnames(3))...
       Dataset.(varnames(4)) Dataset.(varnames(5))...
       Dataset.(varnames(6)) Dataset.(varnames(7))...
       Dataset.(varnames(8)) Dataset.(varnames(9))...
       Dataset.(varnames(10))];
```

Paso 8- Seleccionar Variables o características

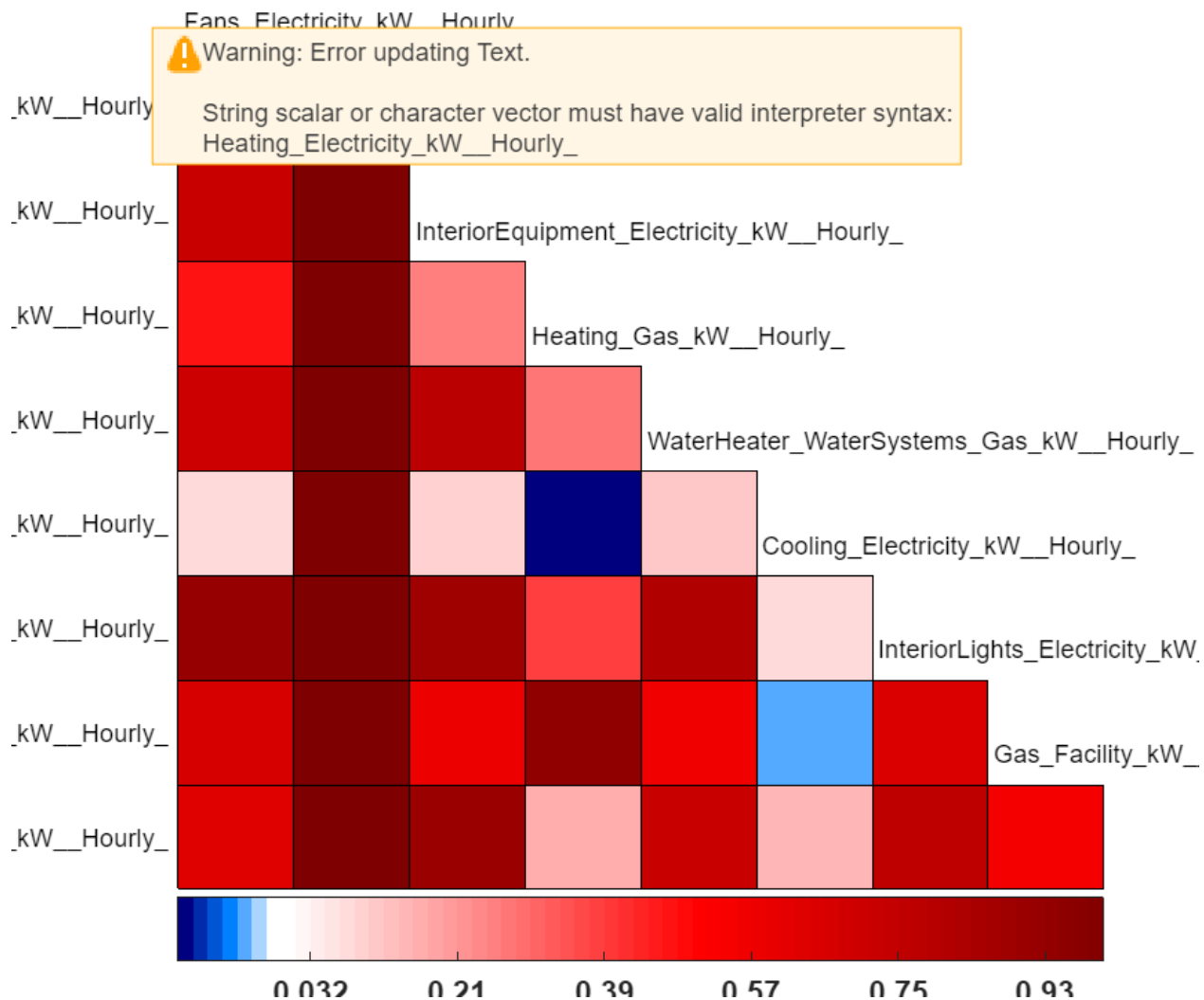
```
%Maximum correlation value allowed
% Default 0.75
threshold = 0.75;

% El numero de variables a analizar debe ser igual al numero de nombres de
% variables
Features_labels=cellstr(varnames(2:end)');
%corrcoef(input)

% Example:
% a=[1:10];b=a+3;c=a.*b;
% corrcoef([a b c])
% 1.0000    1.0000    0.9816
% 1.0000    1.0000    0.9816
% 0.9816    0.9816    1.0000
```

```
[NewDataFeatures,NewFeaturesLabels,LabelsRemove] = Feature_Selection(input,Features_labels,threshold)
```

Electrical Consumption Parameters



```
NewDataFeatures = 8760x4
```

```
3.5862    0    0  136.5859
    0    0    0   3.3599
    0    0    0   3.3599
    0    0    0   3.9319
    0    0    0   3.3599
3.5862  0.0132    0  130.5649
3.5862  0.0074    0  140.4355
3.5862  0.0075    0  135.4560
3.5862    0    0  145.5751
3.5862    0    0  147.6959
    ⋮
```

```
NewFeaturesLabels = 1x4 cell
```

```
'Fans_Electricity_kW_Hourly_' 'Cooling_Electricity_kW_Hourly_' 'Heating_Electric ...
```

```
LabelsRemove = 1x5 cell
```

```
'InteriorLights_Electricity_kW_Hourly_' 'InteriorEquipment_Electricity_kW_Hourly ...
```

```
Warning: Error updating Text.
```

```
String scalar or character vector must have valid interpreter syntax:  
Gas_Facility_kW_Hourly_
```

%Variables eliminadas por superar el threshold


```
LabelsRemove'
```

```
ans = 5x1 cell  
'InteriorLights_Electricity_kW__Hourly_'  
'InteriorEquipment_Electricity_kW__Hourly_'  
'Heating_Gas_kW__Hourly_'  
'InteriorEquipment_Gas_kW__Hourly_'  
'WaterHeater_WaterSystems_Gas_kW__Hourly_'
```

```
%variables que se quedan por no superan el threshold en coreelacion  
NewFeaturesLabels'
```

```
ans = 4x1 cell  
'Fans_Electricity_kW__Hourly_'  
'Cooling_Electricity_kW__Hourly_'  
'Heating_Electricity_kW__Hourly_'  
'Gas_Facility_kW__Hourly_'
```

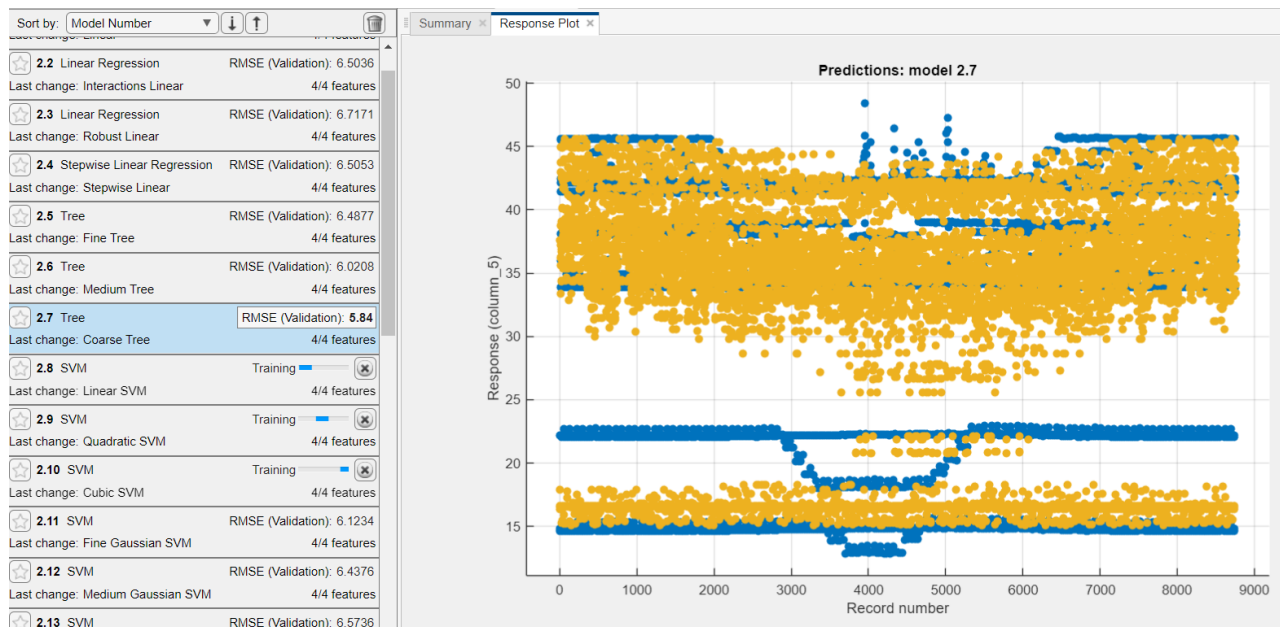
```
ejemplo=[input(1:end-1,4:5) output(2:end,1)]
```

```
ejemplo = 8759x3  
    4.5899    8.1892   14.6498  
    1.5300    7.4902   14.6696  
    1.5300    7.4902   14.6778  
    1.5300    7.4902   14.8248  
    1.5300    7.4902   22.1826  
    4.5899    8.1892   38.1318  
    9.1799   19.4245   45.5973  
    9.1799   26.9147   45.6036  
    9.1799   26.9147   38.1138  
    9.1799   19.4245   42.4352  
    ⋮  
    ⋮  
    ⋮
```

```
clear threshold Features_labels;
```

Paso 9- Seleccionar el algortimo de ML con un menos error de prediccion empleando el toolbox de Matlab Regression Learner

```
% Concatenando la variables de entrada actual con la salida en la hora  
% siguiente  
DataRegression=[NewDataFeatures(1:end-1,:) output(2:end,1)];  
regressionLearner
```



Last change: Medium Neural Network		4/4 features
2.22	Neural Network	RMSE (Validation): 5.8006
Last change: Wide Neural Network		4/4 features
2.23	Neural Network	RMSE (Validation): 5.8089

Model 2.22: Neural Network

Status: Trained

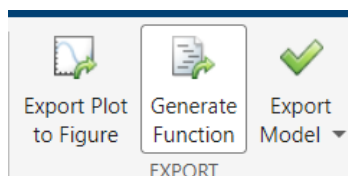
Training Results

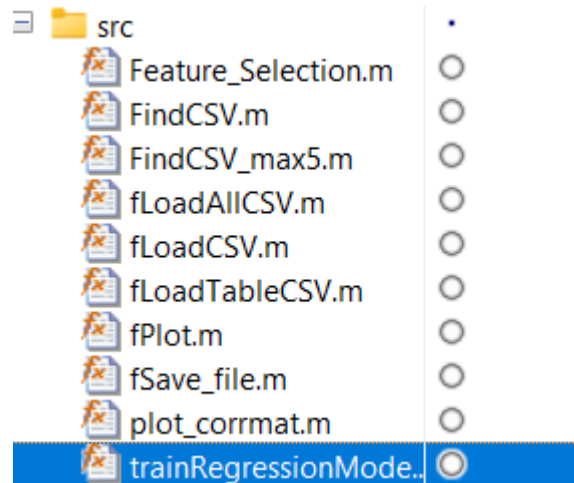
RMSE (Validation)	5.799
R-Squared (Validation)	0.67
MSE (Validation)	33.628
MAE (Validation)	3.9261
Prediction speed	~140000 obs/sec
Training time	372.03 sec

Model Hyperparameters

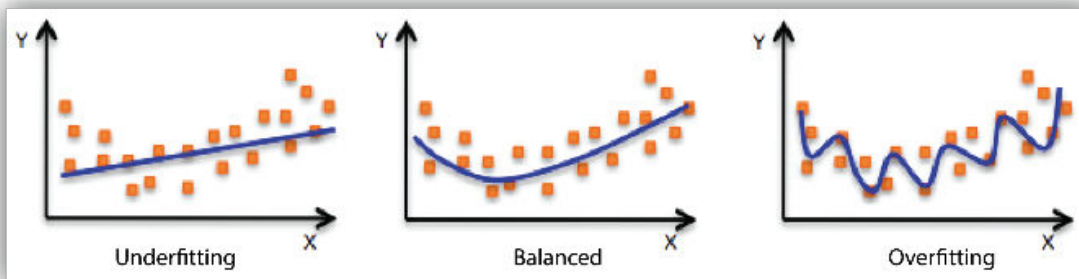
Preset: Wide Neural Network
 Number of fully connected layers: 1
 First layer size: 100
 Activation: ReLU
 Iteration limit: 1000
 Regularization strength (Lambda): 0
 Standardize data: Yes

- Feature Selection: 4/4 individual features selected
- PCA: Disabled
- Optimizer: Not applicable





Paso 10- Dividir el dataset en 70% para entrenar y 30% validar

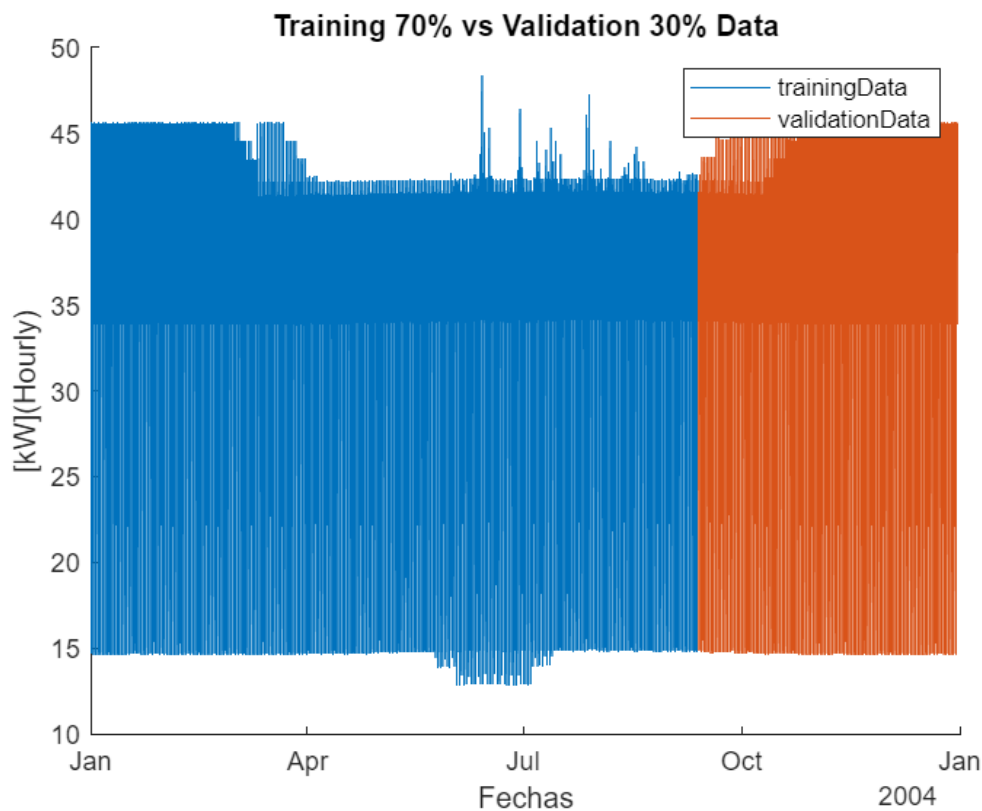


```
%Set de % de entrenamiento
% Dafault 0.70
PSplit = 0.7;
```

```
% Dataset de entrenamiento 70%
trainingData=DataRegression(1:round(end*PSplit),:);
traninTime=time(1:round(end*PSplit)-1);
% Dataset de validcion 30%
validationData=DataRegression(round(end*PSplit)+1:end,:);
validationTime=time(round(end*PSplit)+1:end);

figure;
% Dataset
hold on
plot(traninTime,trainingData(:,end))
plot(validationTime,validationData(:,end))
title(['Training ' num2str(PSplit*100) '% vs Validation '...
      num2str(100-PSplit*100) '% Data']);
xlabel('Fechas')
ylabel('[kW](Hourly)')
```

```
legend('trainingData','validationData')
```



Paso 11- Usando el algoritmo de ML se entrena el modelo de regression (costo computacional)



```
%Esta funcion permite generar un modelo actualizado cada vez que se ejecuta  
% Siempre que el numero de variables de entrada sea la misma y la cantidad  
% de nuevos datos no sea muy alta  
[trainedModel, validationRMSE] = trainRegressionModel(trainingData);  
%Este es el error de entrenaamiento  
validationRMSE
```

```
validationRMSE = 5.9489
```

```
%Permite guardar el modelo entrenado que se encuentra en el workspace  
save("trainedModel.mat","trainedModel")
```

Paso 12- Cargar y validar el modelo entrenado

```
%Cargar el modelo entrenado y guardado
load("trainedModel.mat")

%Usar el modelo entrenado para predecir valores de consumo de energia
yest = trainedModel.predictFcn(validationData(:,1:end-1));

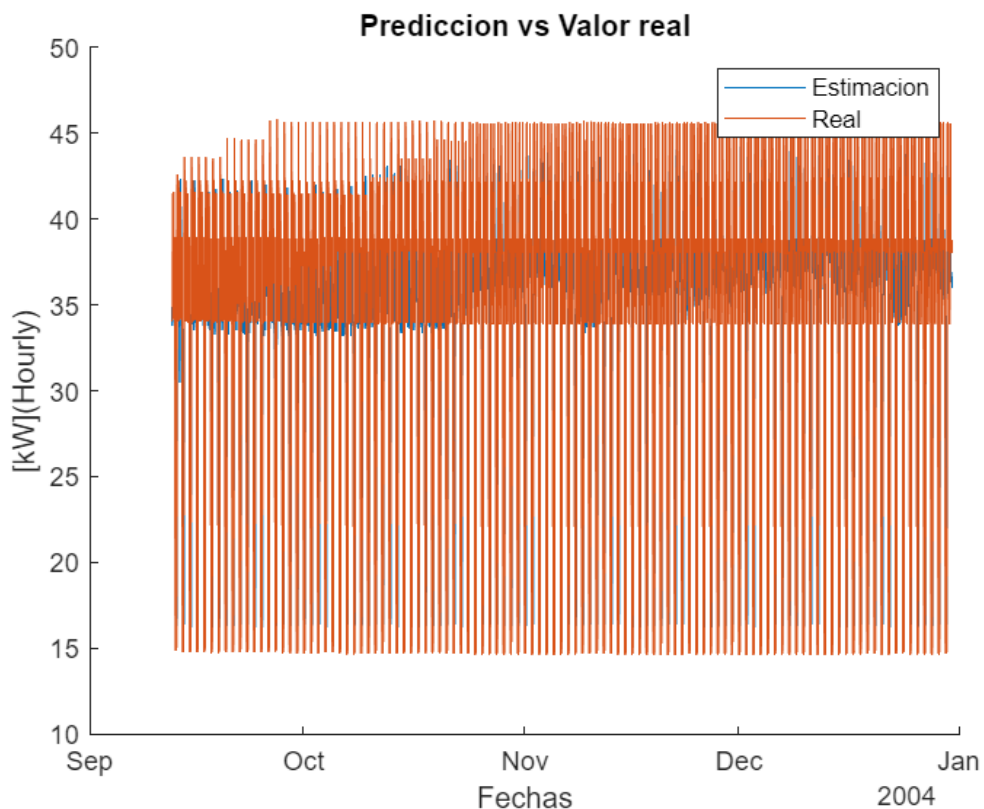
%El valor real de consumo de energia para comparar
yout = validationData(:,end);

%Error de prediccion con datos de validacion
validationRMSE = sqrt(mean((yest - yout).^2))
```

```
validationRMSE = 6.1812
```

paso 13 - Graficar el valor predecido vs el valor real

```
figure;
% Dataset
hold on
plot(validationTime,yest)
plot(validationTime,yout)
title('Prediccion vs Valor real');
xlabel('Fechas')
ylabel('[kW](Hourly)')
legend('Estimacion','Real')
```



paso 14 - Mejorar el modelo de prediccion

