226359

September 30, 2019

1 Jimi Togni - RA: 226359

1.1 Parte 1 - Atividade teorica

Atividades Teóricas

EFC 1

Jimi Togni RA-226359

$\boxed{\text{Ex } 1}$ a) $P(X=1) = \frac{1}{8} + \frac{1}{3} = \frac{11}{24}$

$P(Y=1) = \frac{3}{8} + \frac{1}{3} = \frac{17}{24}$

$P(X=0) = \frac{1}{6} + \frac{3}{8} = \frac{13}{24}$

$P(Y=0) = \frac{1}{6} + \frac{1}{8} = \frac{7}{24}$

b) $P(X=0 \mid Y=0) = \dfrac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{8}} \simeq 57{,}14\%$

c) $E(X) = \frac{1}{8} + \frac{1}{3} = \frac{11}{24}$

$E(Y) = \frac{3}{8} + \frac{1}{3} = \frac{17}{24}$

$\boxed{Ex\ 1}$ $\boxed{d)}$ NÃO são independentes, uma vez que

$$P(X=1 \mid Y=1) = \frac{\frac{1}{8}}{\frac{3}{8}+\frac{1}{3}} \simeq 0{,}176 \neq 0{,}125$$

$\boxed{Ex\ 2}$ $\boxed{a)}$ $P(X=1) = \frac{3}{4}$ $\quad$ $H = -\sum_{i=0}^{n} p_i \log_2 p_i$

$$H(X) = \left(-\frac{3}{4}\log_2\frac{3}{4}\right) + \left(-\frac{1}{4}\log_2\frac{1}{4}\right) \triangleq 0{,}811$$

$$H(Y) = \left(-\frac{3}{8}\log_2\frac{3}{8}\right) + \left(-\frac{5}{8}\log_2\frac{5}{8}\right) \triangleq 0{,}954$$

$$H(X,Y) = \left(-\frac{1}{4}\log_2\frac{1}{4}\right) + \left(-\frac{3}{8}\log_2\frac{3}{4}\right) + \left(-\frac{3}{8}\log_2\frac{3}{4}\right) = 1{,}561$$

$\boxed{b)}$ $H(Y|X) = \left(-\frac{1}{4}\log_2 1\right) + \left(-\frac{3}{8}\log_2 0{,}5\right) + \left(-\frac{3}{8}\log_2 0{,}5\right) = 0{,}75$

$H(X|Y) = \left(-\frac{1}{4}\log_2 0{,}4\right) + \left(-\frac{3}{8}\log_2 1\right) + \left(-\frac{3}{8}\log_2 0{,}6\right) = 0{,}60$

$\boxed{c)}$ $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \triangleq$

$\triangleq 0{,}204$

$\boxed{Ex\ 3}$ $\boxed{a)}$

$$g(x) = g_1(x) - g_2(x)$$

$$g(x) = 0 = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x+1}{1}\right)^2} - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-1}{1}\right)^2}$$

$$0 = \frac{1}{\sqrt{2\pi}} \cdot \left(e^{-\frac{1}{2}\left(\frac{x+1}{1}\right)^2} - e^{-\frac{1}{2}\left(\frac{x-1}{1}\right)^2} \right)$$

$$e^{-\frac{1}{2}\left(\frac{x+1}{1}\right)^2} = e^{-\frac{1}{2}\left(\frac{x-1}{1}\right)^2} \Rightarrow$$

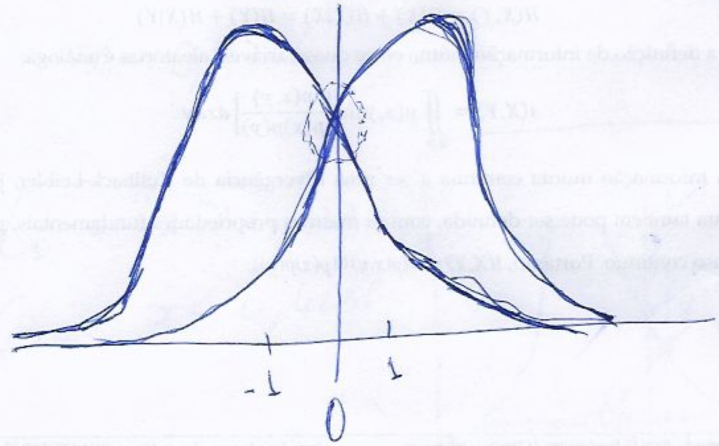
$$\Rightarrow -\frac{1}{2}(x+1)^2 = -\frac{1}{2}(x-1)^2 =$$

$$\Rightarrow -\frac{1}{2}(x^2 2x + 1) = -\frac{1}{2}(x^2 - 2x + 1)$$

$$x = 0$$

$\therefore C_1 = x > 0$

$C_2 = x < 0$



$\boxed{Ex\ 3}$ $\boxed{b)}$ $g(x) = 0{,}\underbrace{7(x)}_{g_1} - 0{,}\underbrace{3(x)}_{g_2}$

$$g(x) = 0 = \frac{1}{\sqrt{2\pi}}\, 0{,}7e^{-\frac{1}{2}\left(\frac{x+1}{1}\right)^2} - \frac{1}{\sqrt{2\pi}}\, 0{,}3e^{-\frac{1}{2}\left(\frac{x-1}{1}\right)^2}$$

$$\Rightarrow \frac{1}{\sqrt{2\pi}}\left(0{,}7e^{-\frac{1}{2}\left(\frac{x+1}{1}\right)^2} - 0{,}3e^{-\frac{1}{2}\left(\frac{x-1}{1}\right)^2}\right)$$

$$\Rightarrow 0{,}7^{-\frac{1}{2}\left(\frac{x+1}{1}\right)^2} = 0{,}3^{-\frac{1}{2}\left(\frac{x-1}{1}\right)^2} =$$

$$\Rightarrow \ln 0{,}7 + \left(\frac{-1}{2}\right)(x+1)^2 = \ln 0{,}3 + \left(\frac{-1}{2}\right)(x-1)^2 =$$

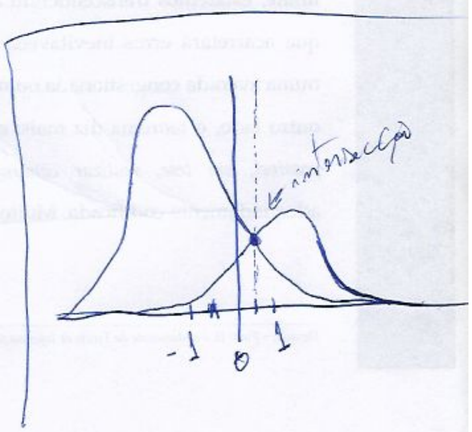
$$\Rightarrow \quad -2x = \ln 0{,}3 - \ln 0{,}7 \quad =$$

$$\Rightarrow \quad x = \frac{\ln 0{,}3 - \ln 0{,}7}{2} =$$

$$\Rightarrow \quad x = 0{,}4236$$

$$\therefore\ C_1 = x > 0{,}4236$$

$$C_2 = x < 0{,}4236$$



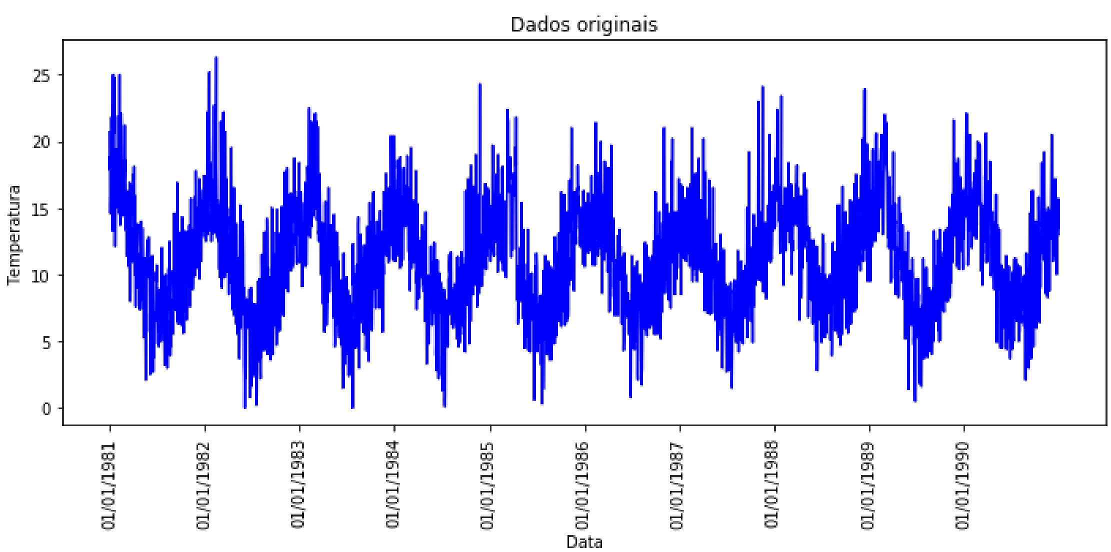
1.2 Parte 2 – Atividade computacional

Ultimos 10 registros da base de dados.

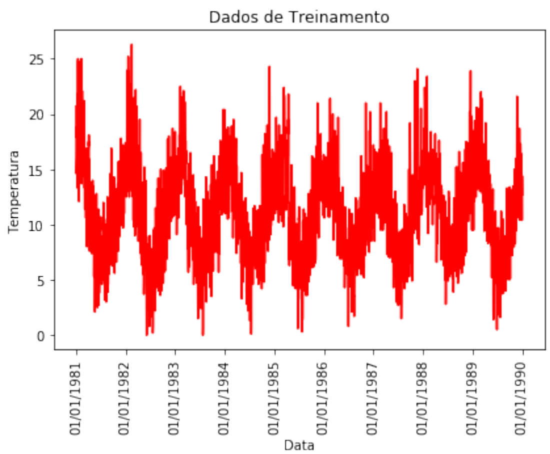
```
           Data  Temperature
3640  22/12/1990         13.2
3641  23/12/1990         13.9
3642  24/12/1990         10.0
3643  25/12/1990         12.9
3644  26/12/1990         14.6
3645  27/12/1990         14.0
3646  28/12/1990         13.6
3647  29/12/1990         13.5
3648  30/12/1990         15.7
3649  31/12/1990         13.0
```

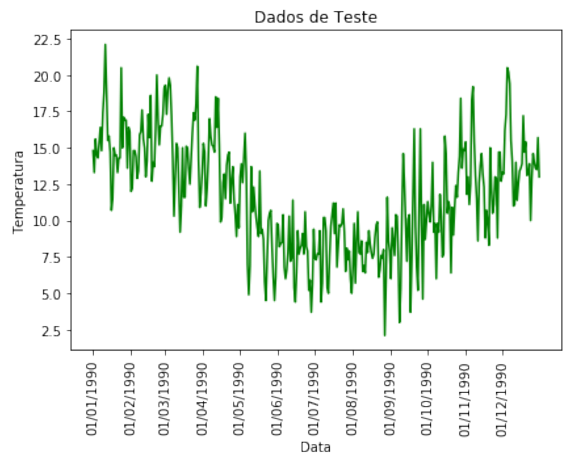
```
/usr/lib/python3/dist-packages/ipykernel_launcher.py:7: UserWarning: This pattern has match gro
  import sys
```

Grafico de toda a base de dados.

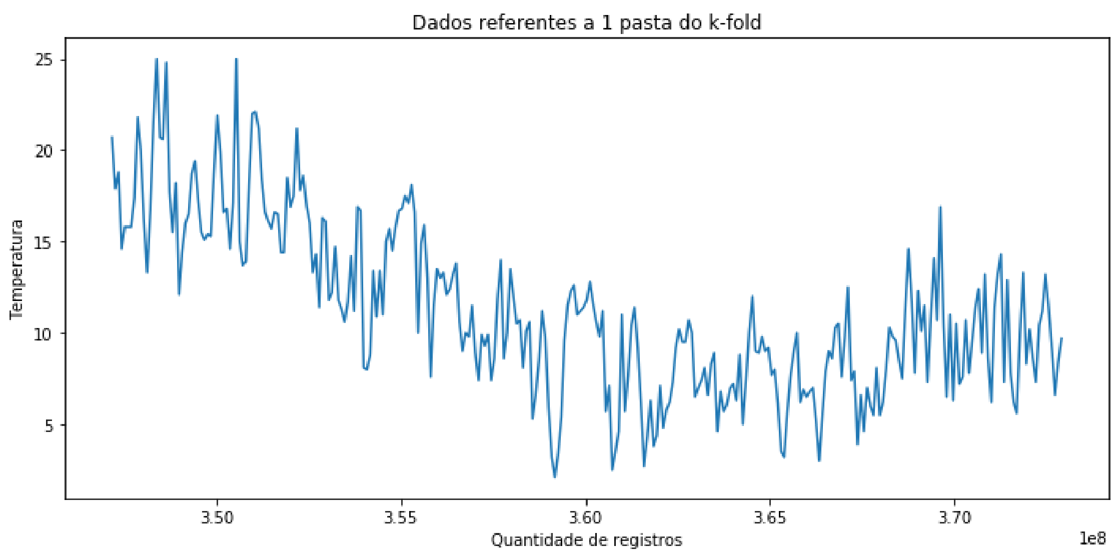


Divisão dos dados para treinamento e, o ultimo ano, para teste.





Para evitar sobreajuste, foi utilizada a tecnica de k-fold, que resumidamente, consiste em de divisoes do dataset de treinamento e validacao em partes menores, o grafico abaixo demonstra a distribuicao para uma das 10 pastas

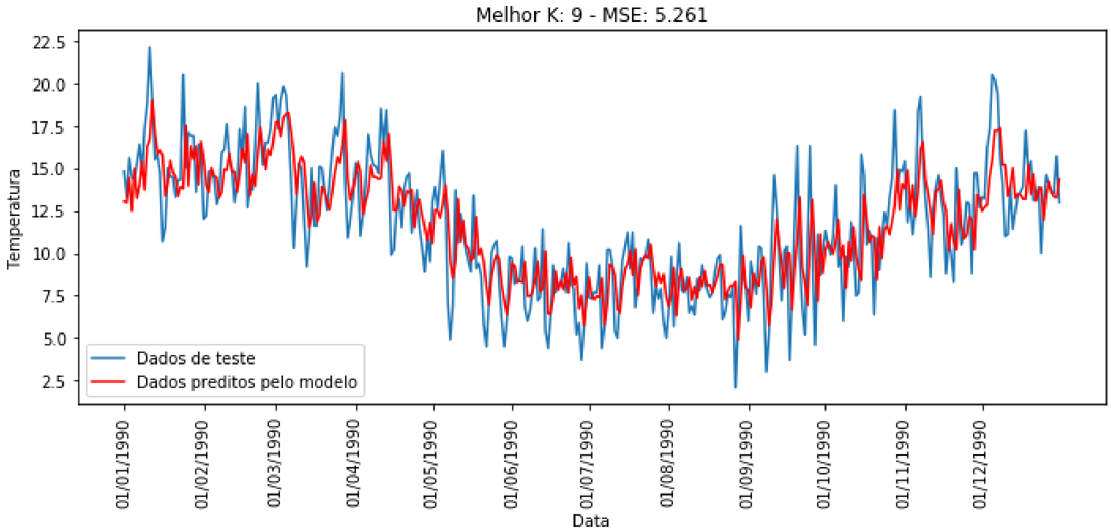


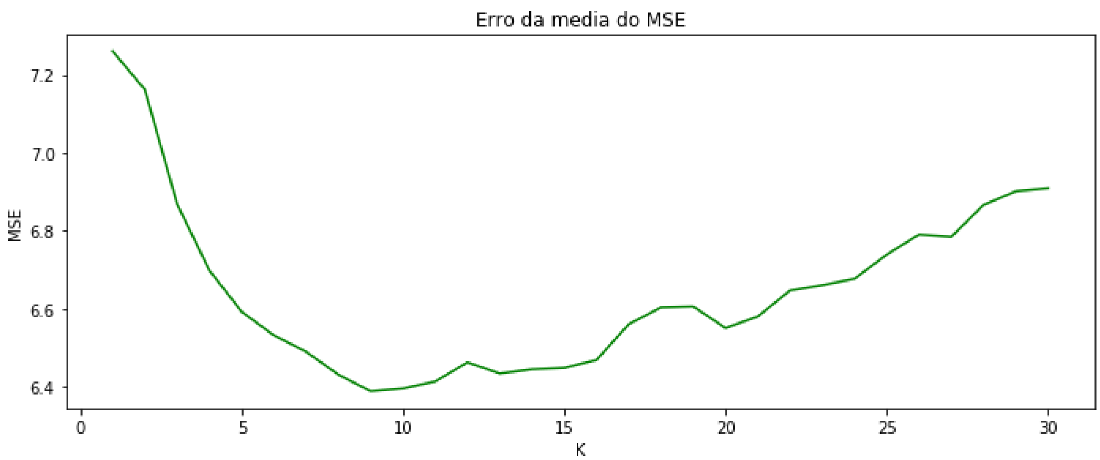
Exercício 1 Estipulando o range para os `k` atrasos

```
Valores:
Min: 0.0, Max: 24.72
```

Executando a Regressão Linear, com K variando de 1 a 30 e k-fold variando de 1 a 20 folds
Resultados:

```
Melhores valores encontrados:
K:  9
k-fold:  6
```





No primeiro gráfico, pode-se observar o resultado do modelo, utilizando para o hiper-parâmetro k = 9, o erro médio com MSE = 5.261.

No segundo gráfico é demonstrada a variação do MSE ao longo das 30 pastas.

Demonstra-se na tabela a baixo, 10 resultados apos as iteracoes.

	K	k-fold	Fold de Validacao	Média MSE
272	16	3	3	6.105617
254	15	3	3	6.113649
236	14	3	3	6.118070
218	13	3	3	6.130893
149	9	6	6	6.134120
167	10	6	6	6.134760

```
164  10     3             3   6.143207
146   9     3             3   6.144946
182  11     3             3   6.145171
166  10     5             5   6.150721
```

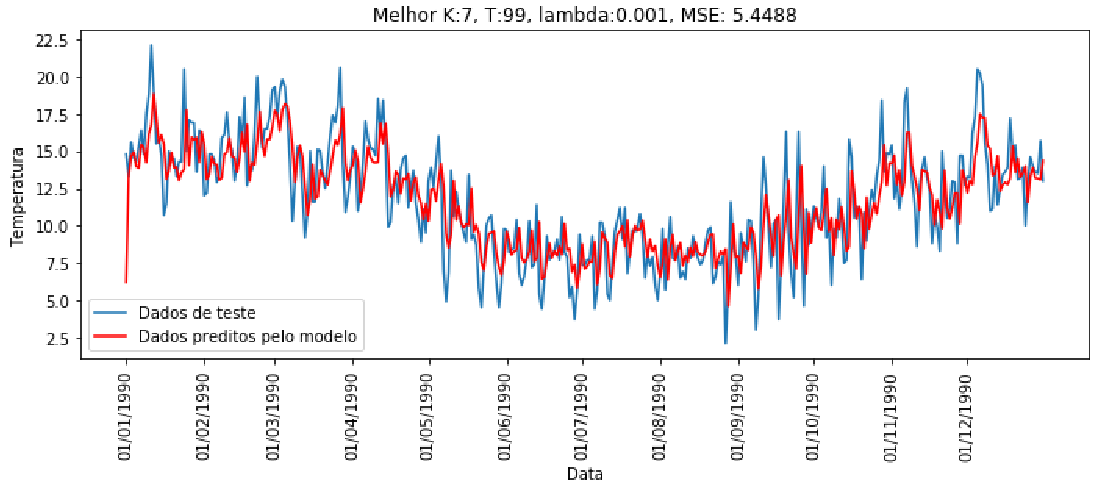
Exercício 2 **Para a quantidade de atributos gerados, vamos utilizar 100 iterações, com uma distribuição uniforme que variou de -1 até 1.**

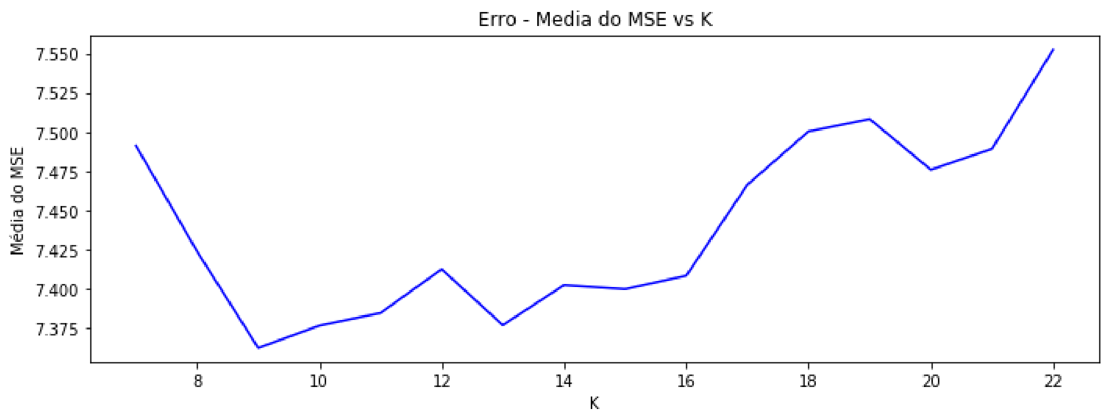
OS valores do lambda variaram de 1e+1 - 1e-4

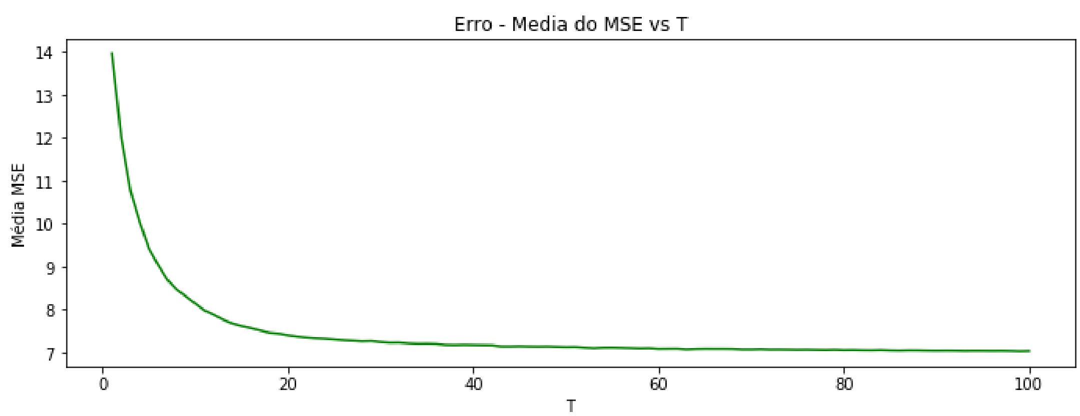
Para a normalização, os dados foram enquadrados entre os valores mínimo e máximo, obtidos anteriormente.

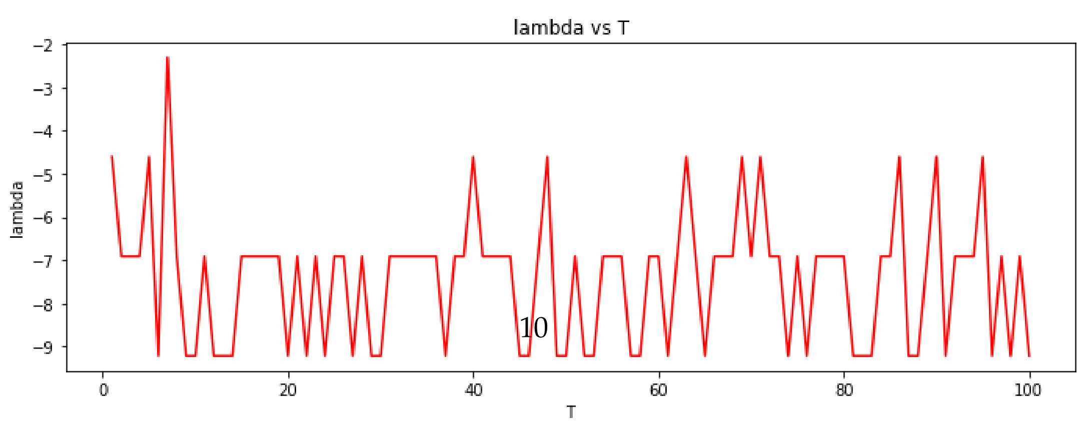
Os valores de K estão entre 7 a 22, e o K-Fold utilizado foi de 1 até 10 folds.

```
6 valores para o lambda -> [1.e+01 1.e+00 1.e-01 1.e-02 1.e-03 1.e-04]


1 -> K: 7
2 -> K: 8
3 -> K: 9
4 -> K: 10
5 -> K: 11
6 -> K: 12
7 -> K: 13
8 -> K: 14
9 -> K: 15
10 -> K: 16
11 -> K: 17
12 -> K: 18
13 -> K: 19
14 -> K: 20
15 -> K: 21
16 -> K: 22


Melhores resultados obtidos:
k-fold:  1
K'atrasos':  7
T:  99
lambda:  0.001
MSE:  6.852442664258398
```


Melhor K:7, T:99, lambda:0.001, MSE: 5.4488
Temperatura
Data
Dados de teste
Dados preditos pelo modelo


Erro - Media do MSE vs K
Média do MSE
K


Erro - Media do MSE vs T
Média MSE
T


lambda vs T
lambda
T

Os dados obtidos após a passagem pelo segundo modelo de regressão linear foram:

	K	K-Fold	Fold de Validacao	T	Regularizacao	Média MSE
11204	16	3	2	95	0.0100	6.533775
11205	16	3	2	96	0.0001	6.541001
7566	13	3	2	93	0.0010	6.575138
8770	14	3	2	85	0.0010	6.575419
7562	13	3	2	89	0.0010	6.578258
9974	15	3	2	77	0.0010	6.581564
11201	16	3	2	92	0.0010	6.581609
8776	14	3	2	91	0.0001	6.583636
11185	16	3	2	76	0.0001	6.584036
11200	16	3	2	91	0.0010	6.584299

Observa-se que, nos dois modelos implementados nesta lista de exercícios, ambos apresentaram resultados aproximados em relação ao erro médio e a predição dos dados, porém, o ganho de acurácia no segundo modelo, foi visivelmente melhor