

# Exercícios de Fixação de Conceitos 1 - EFC1 - IA048

Marcelo Eduardo Pederiva RA: 122580

## Parte 1 - Atividades Teóricas

### Exercício 1

a)

$$P(X) = \sum_x P(X = x)$$

$$\begin{aligned} P(X = 0) &= P(X = 0, Y = 0) + P(X = 0, Y = 1) \\ P(X = 0) &= 0.5 + 0.05 \end{aligned}$$

$$\boxed{P(X = 0) = 0.55}$$

$$\begin{aligned} P(X = 1) &= P(X = 1, Y = 0) + P(X = 1, Y = 1) \\ P(X) &= 0.3 + 0.15 \end{aligned}$$

$$\boxed{P(X = 1) = 0.45}$$

$$P(Y) = \sum_y P(Y = y)$$

$$\begin{aligned} P(Y = 0) &= P(Y = 0, X = 0) + P(Y = 0, X = 1) \\ P(Y = 0) &= 0.5 + 0.3 \end{aligned}$$

$$\boxed{P(Y = 0) = 0.8}$$

$$\begin{aligned} P(Y = 1) &= P(Y = 1, X = 0) + P(Y = 1, X = 1) \\ P(Y = 1) &= 0.05 + 0.15 \end{aligned}$$

$$\boxed{P(Y = 1) = 0.2}$$

b)

$$P(X = 1|Y = 1) = \frac{P(Y = 1, X = 1)}{P(Y = 1)}$$
$$P(X = 1|Y = 1) = \frac{0.15}{0.2}$$

$$\boxed{P(X = 1|Y = 1) = 0.75}$$

c)

Para observar se as variáveis são descorrelacionadas vamos calcular a  $cov(X, Y)$ . Se  $cov(X, Y) = 0$ , temos que as variáveis são descorrelacionadas.

$$cov(X, y) = E[(X - \mu_x) - (Y - \mu_y)]$$
$$= \sum_{x,y} (x - \mu_x) - (y - \mu_y) P(X = x, Y = y)$$

$$\mu_x = 0 * 0,5 + 0 * 0,05 + 1 * 0,3 + 1 * 0,15 = 0,45$$

$$\mu_y = 0 * 0,5 + 0 * 0,3 + 1 * 0,05 + 1 * 0,15 = 0,20$$

Dessa forma temos,

$$cov(X, Y) = (0 - 0,45)(0 - 0,2)0,5 + (0 - 0,45)(1 - 0,2)0,5$$
$$+ (1 - 0,45)(0 - 0,2)0,5 + (1 - 0,45)(1 - 0,2)0,5$$
$$cov(X, Y) = 0,045 - 0,018 - 0,033 + 0,66$$
$$cov(X, Y) = 0,06 \neq 0$$

Assim temos que as variáveis são correlacionadas.

d)

Temos que, para uma variável ser independente, ela deve satisfazer as seguintes condições:

- $P(X|Y) = P(X)$  para todos valores de X e Y.
- $P(X \cap Y) = P(X) * P(Y)$  para todos valores de X e Y.

Através dos resultados obtidos pelo item a) e b) temos que,

$$P(X = 1|Y = 1) = 0.75 \neq 0.45 = P(X = 1)$$

Sendo assim podemos admitir que as duas variáveis não são independentes.

e)

- $H(X)$

$$\begin{aligned} H(X) &= - \sum_x P(X = x) \log_2[P(X = x)] \\ &= -(P(X = 0) \log_2[P(X = 0)] + P(X = 1) \log_2[P(X = 1)]) \\ &= 0,55(-0,8625) + 0,45(-1,1520) \end{aligned}$$

$$\boxed{H(X) = 0,9928}$$

- $H(Y)$

$$\begin{aligned} H(Y) &= - \sum_y P(Y = y) \log_2[P(Y = y)] \\ &= -(P(Y = 0) \log_2[P(Y = 0)] + P(Y = 1) \log_2[P(Y = 1)]) \\ &= -(0,8(-0,3219) + 0,2(-2,3219)) \end{aligned}$$

$$\boxed{H(Y) = 0,7219}$$

- $H(X, Y)$

$$\begin{aligned} H(X, Y) &= - \sum_x \sum_y P(X = x, Y = y) \log_2[P(X = x, Y = y)] \\ &= -(P(X = 0, Y = 0) \log_2[P(X = 0, Y = 0)] + P(X = 0, Y = 1) \log_2[P(X = 0, Y = 1)] \\ &\quad + P(X = 1, Y = 0) \log_2[P(X = 1, Y = 0)] + P(X = 1, Y = 1) \log_2[P(X = 1, Y = 1)]) \\ H(X, Y) &= -(0,5 \log_2(0,5) + 0,05 \log_2(0,05) + 0,3 \log_2(0,3) + 0,15 \log_2(0,15)) \end{aligned}$$

$$\boxed{H(X, Y) = 1,6477}$$

- $H(X|Y)$

$$H(X, Y) = H(Y) + H(X|Y)$$

$$H(X|Y) = H(X, Y) - H(Y)$$

$$H(X|Y) = 1,6477 - 0,7219$$

$$\boxed{H(X|Y) = 0,9258}$$

- $H(Y|X)$

$$H(X, Y) = H(X) + H(Y|X)$$

$$H(Y|X) = H(X, Y) - H(X)$$

$$H(Y|X) = 1,6477 - 0,9928$$

$$\boxed{H(Y|X) = 0,6549}$$

f)

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$I(X, Y) = 0,9928 - 0,9258$$

$$\boxed{I(X, Y) = 0,067}$$

## Parte 2 - Atividade Computacional

### Exercício 1

Neste primeiro exercício, utilizamos um aproximador linear para prever o valor médio de manchas solares nos próximos meses. Para isso, foi separado os últimos dez anos de dados (2010-2019) como conjunto de teste. Da mesma forma, os dados restantes foram usados como conjunto de treinamento.

Na etapa de treinamento, foi utilizado o método de validação cruzada. Para isso, determinando 20% do conjunto de treinamento como dados de validação, foi utilizado 5 "pastas" para realizar a validação cruzada do treinamento.

#### 1.

Nesta etapa, observamos a influência no erro da predição quando aumentamos a quantidade de meses anteriores utilizados no treinamento.

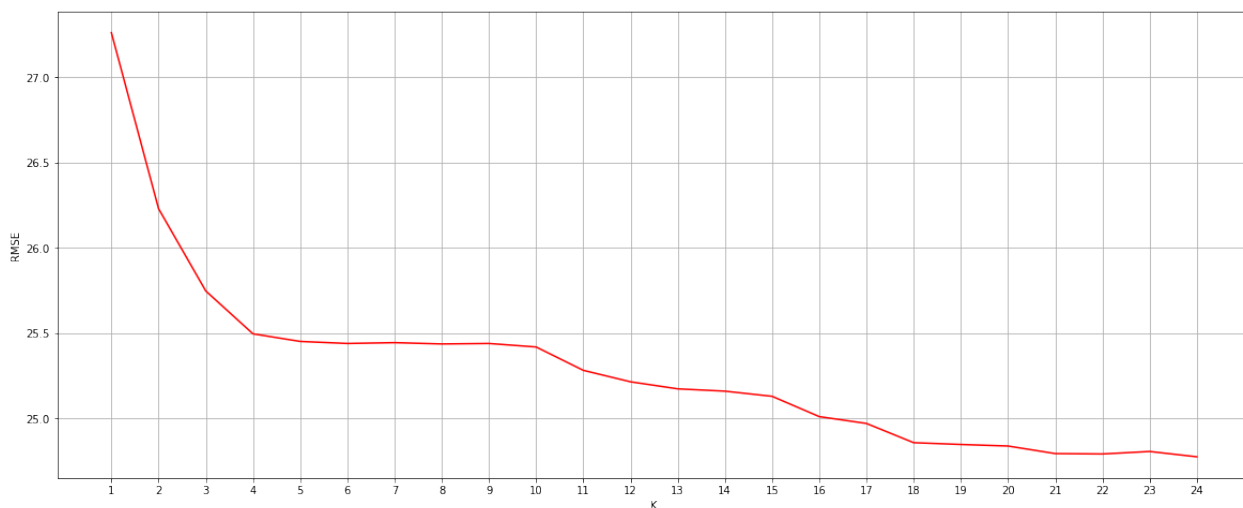


Figura 1: Relação do erro na predição com o valor K.

Como podemos observar na Figura 1, o aumento de meses anteriores nos dados teve uma influência positiva na predição do valor médio de manchas solares no próximo mês.

Neste treinamento o menor valor médio da raiz quadrada do erro quadrático médio (RMSE, do inglês *root mean squared error*) da predição alcançou um valor de **RMSE=24.7761** com **K = 24**.

## 2.

Dessa forma, utilizando a matriz  $W$  treinada com o  $K=24$ , prevemos os valores do conjunto de teste e comparamos com os valores reais esperados (Figura 2).

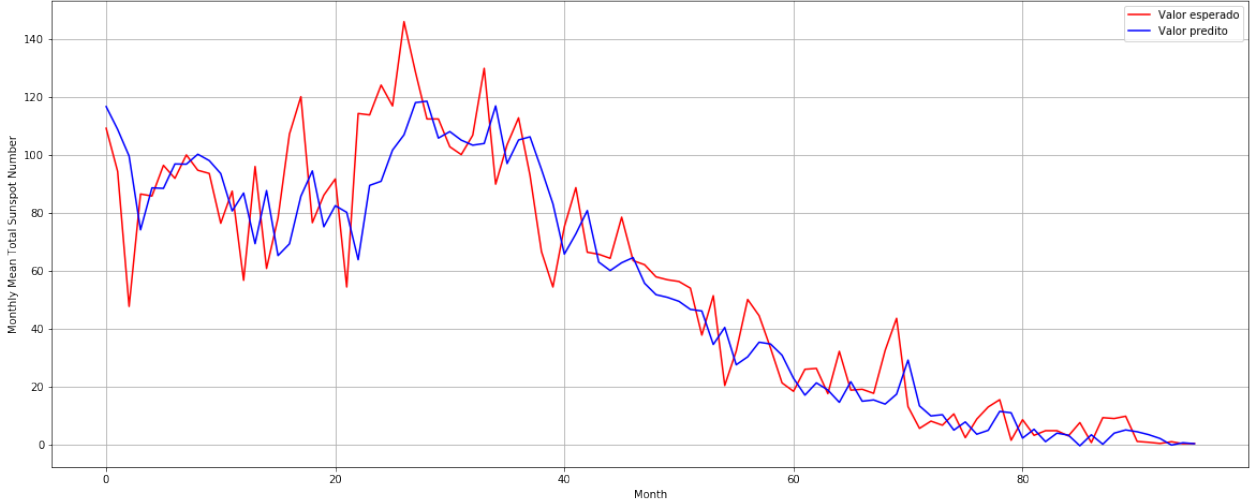


Figura 2: Comparação do valor predito com o valor esperado.

Podemos observar, que a curva predita tem uma boa aproximação da curva esperada, entretanto, possui um deslocamento a direita dos dados reais. Isto pode ocorrer pelo preditor linear valorizar mais os meses mais próximos da predição, ou seja, os meses mais recentes recebem um peso maior na matriz  $W$ . Dessa forma, a predição sempre propõe um valor próximo ao apresentado no mês anterior.

Por fim, o Preditor Linear alcançou um valor de **RMSE = 15.8979** no conjunto de teste.

## Exercício 2

Neste exercício, implementamos uma aproximação não-linear para resolução do mesmo problema. Para isso, os dados que utilizamos para treinar o modelo sofreu um pré-processamento, uma normalização, seguido por uma multiplicação com uma matriz com números uniformemente aleatórios ( $W_k$ ) e posteriormente aplicamos a tangente hiperbólica nesse resultado (Equação 1).

$$x'_k(n) = \tanh(\mathbf{W}_k^T \mathbf{x}_{norm}(n)) \quad (1)$$

Os dados foram normalizados para valores entre 0 e 0,5 e a matriz  $W_k$  foi gerada por números uniformemente aleatórios entre 0 e 0,5.

Para avaliar os resultados da predição não-linear em comparação com da predição linear, foi utilizado o mesmo método de validação cruzada com 5 "pastas".

a)

Tendo como base o  $K=8$ , iremos observar a influência do tamanho da matriz ( $W_k$ ) no erro de predição (RMSE) junto aos dados de validação.

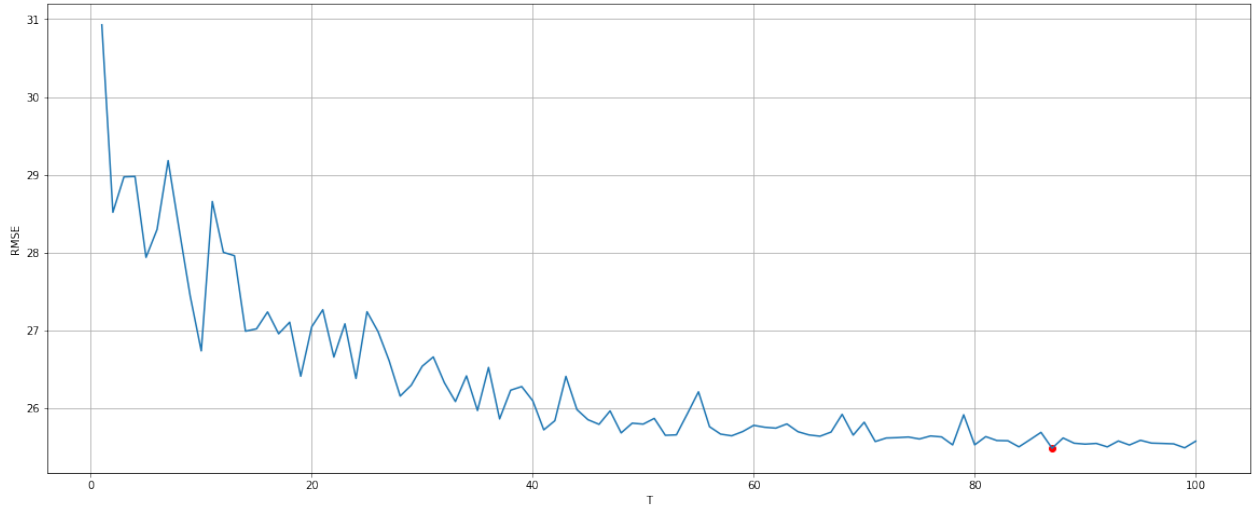


Figura 3: Erro de validação de cada valor de  $T$ , ponto vermelho representa o  $T$  com menor RMSE.

Como podemos observar na Figura 3, o aumento do número de atributos ( $T$ ) proporciona um menor erro nas predições de validação. Sendo assim, com o  $K=8$ , obtivemos o menor **RMSE = 25.4811** com  **$T = 87$** .

b)

A seguir, iremos observar a variação do termo de regularização para cada  $T$ . Este coeficiente foi variado entre  $2^{-10}$  a  $2^{10}$  para cada  $T$ , onde foi destacado o coeficiente que gerava o menor erro de validação. Podemos observar pela Figura 4 o melhor valor do termo de regularização para cada número de atributos.

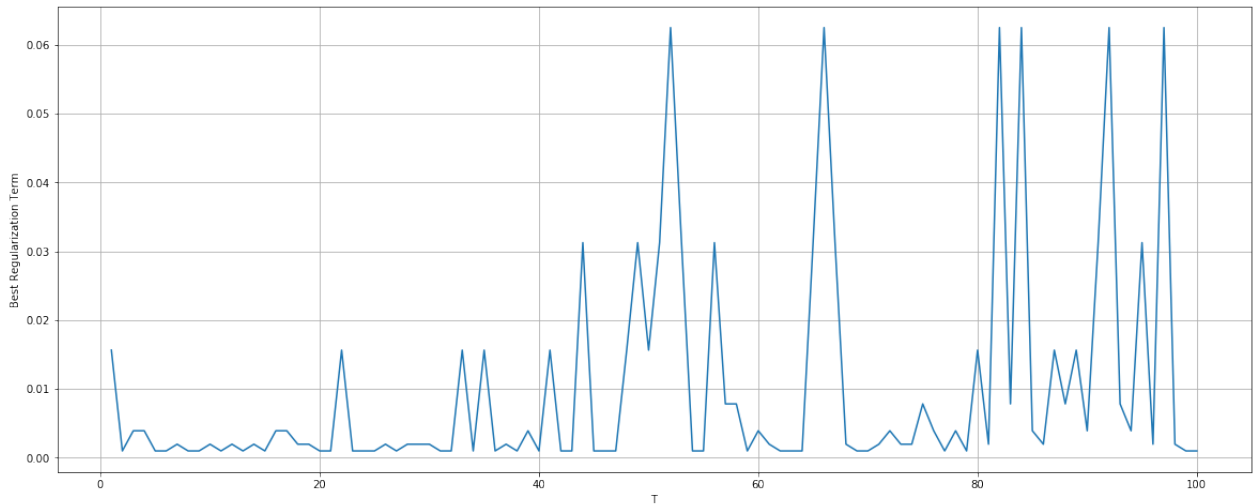


Figura 4: Melhor coeficiente de regularização para cada  $T$ .

Assim, tendo  $T=87$  como melhor candidato proposto pelo item a), encontramos que o melhor coeficiente de regularização para este  $T$  é o valor **0.0078125**.

c)

Por fim, utilizando os melhores resultados dos itens anteriores geramos um estimador não-linear para prever o conjunto de teste.

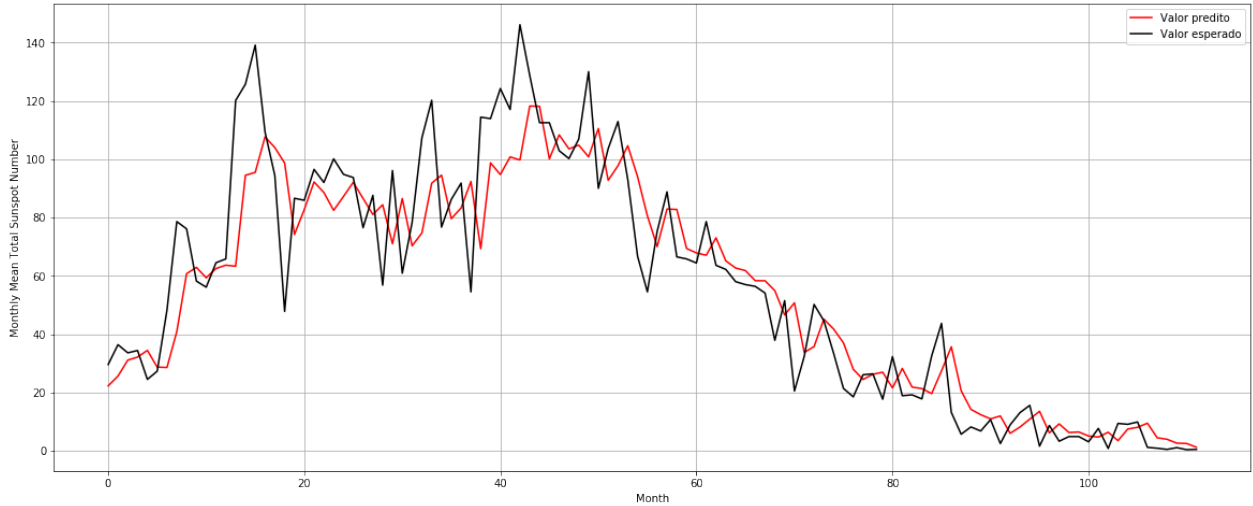


Figura 5: Comparação do valor predito com o valor esperado.

Esta configuração de Ridge Regression resultou em um **RMSE = 16.7270**.



## Conclusão

Neste trabalho utilizei meu RA (122580) como "seed" para gerar os números aleatórios da matriz  $W_k$ , outras "seeds" podem resultar numa pequena variação da predição não-linear. Também foi variado tanto os valores de normalização, como o intervalo de números aleatórios gerados pela matriz  $W_k$ . Os valores definidos no trabalho final foram os que retornaram menores erros junto aos dados de validação.

Por fim, o preditor não-linear apresentou um resultado próximo ao do preditor linear. Comparando os dois métodos com o mesmo K (K=8), observamos o seguinte resultado:

Preditor Linear	RMSE = 16.5682
Preditor Não-Linear	RMSE = 16.7270

Também podemos observar na Figura 6 os valores preditos de cada estimador, preditor linear em azul, preditor não-linear em vermelho e valor esperado em preto.

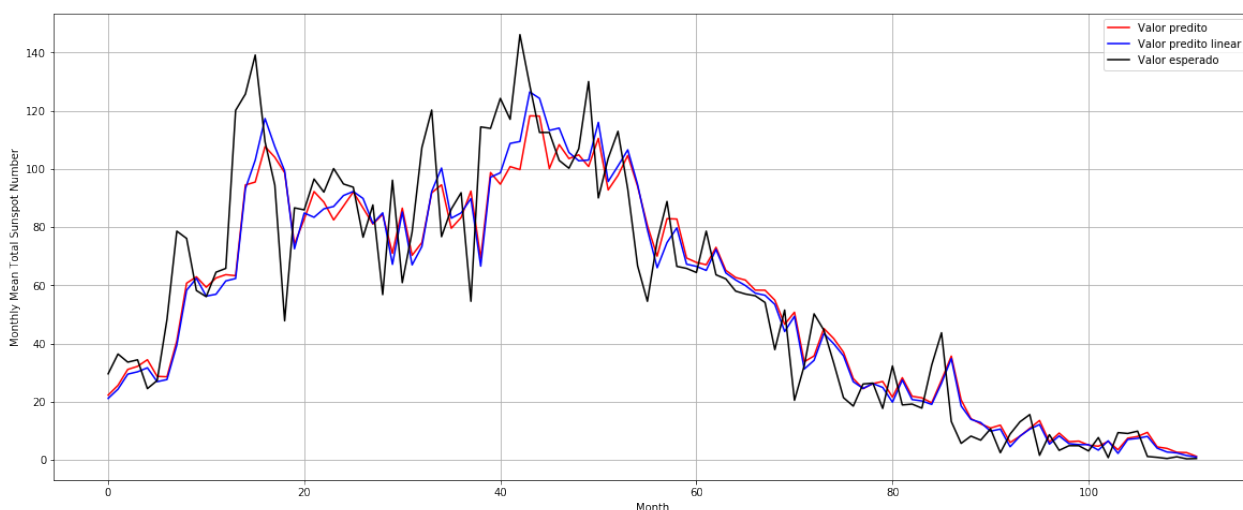


Figura 6: Comparação dos Preditores junto aos dados de Teste, com K=8.

Apesar de ser possível encontrar um resultado levemente superior para o preditor não-linear, com outros números aleatórios gerados pela matriz  $W_k$ , observamos que o preditor linear satisfaz a predição desta serie não-estacionária de forma satisfatória e equivalente à de um preditor não-linear.