

**Tabela de Conteúdos**

1. **Introdução**
2. **Modelação do Problema**
3. **Codificação do Problema**
4. **Resultados Variados**
5. **Conclusão**

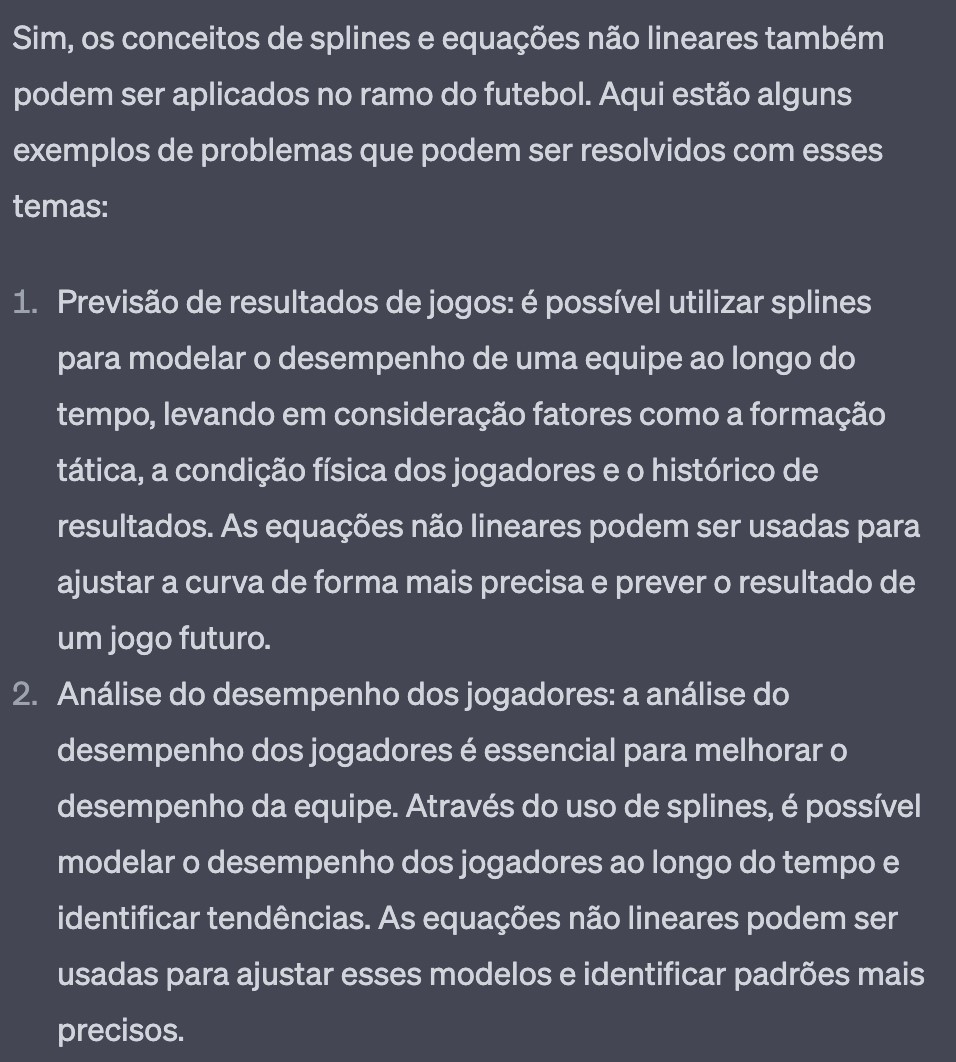
# Introdução

O futebol é amplamente discutido em Portugal, mas muitas vezes as expectativas dos fãs em relação ao desempenho dos jogadores são baseadas apenas em fanatismo.

Deste modo, tendo já definido o tema que pretendíamos falar, perguntamos ao “ChatGPT” como seria possível relacionar o tema com a unidade curricular “Métodos Numéricos e Otimização não linear”, de forma que fossemos capazes de prever estatísticas face aos desempenhos dos jogadores.

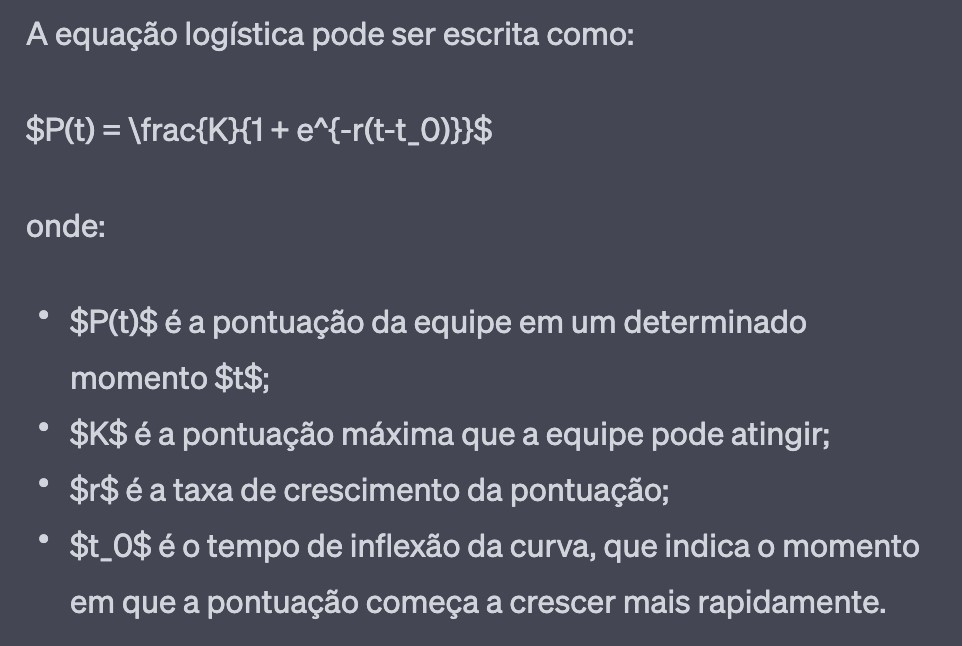
**Pergunta:**

Há algum problema real que se possa resolver com os temas: Splines e Equações não lineares no ramo do futebol?



# Modelação do Problema

A primeira proposta de resolução do “ChatGPT” foi a implementação da equação logística que está na imagem seguinte, no entanto, como nós queríamos, inicialmente, um modelo que relacionasse as “Splines” com a “Fórmula fundamental do erro”, não era de todo uma proposta válida para progredir.



Então, após discutir o tema com o "ChatGPT", decidimos utilizar o modelo que se segue que já combinava os conceitos de "Splines" e da "Fórmula Fundamental do Erro".

**Pergunta:**

Consegues usar uma função spline do MATLAB para prever o número de golos num ponto num intervalo de 30 jogos?

**Resposta:**

Segue um exemplo de um Código:

# MATLAB code

% Define the x and y coordinates of the data points

x = [1 2 3 4 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29]; y = [2 1 1 1 2 3 2 1 3 2 2 2 1 1 1 1 3 2 4 2 1 1 1 1 1 0 1];

% Define the x values where you want to predict the number of goals xx = 1:0.1:5;

% Create the spline yy = spline(x,y,xx);

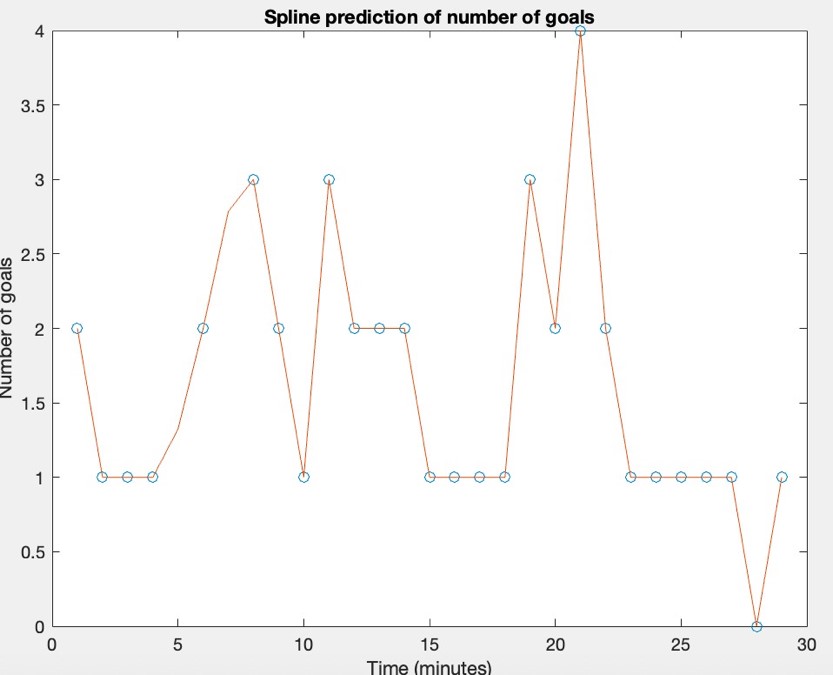
% Plot the results plot(x,y,'o',xx,yy); xlabel('Time (minutes)'); ylabel('Number of goals');

title('Spline prediction of number of goals');

Neste exemplo, primeiro definimos as coordenadas x e y dos pontos que queremos interpolar. Os valores de x representam o tempo e os de y o número de golos de cada equipa. Depois definimos os valores de xx que representam o número de golos onde nos queremos prever o número de golos. Usamos uma função de spline para criar uma spline cúbica que passa pelos pontos e prevê o número de golos nos valores de xx. Finalmente plotamos os valores originais com os que a spline previu.

# -----------------------------------------------------------------------------------------------------------------

No entanto, ao visualizar o resultado obtido, notamos que a curva gerada pela spline não era a melhor previsão possível, uma vez que não correspondia aos resultados esperados. Resultando em um modelo insuficiente que não atingiu nossas expectativas. Logo tivemos que alterar novamente o modelo a utilizar para a resolução do nosso problema inical.



Depois de algumas tentativas em utilizar métodos numéricos e otimização não linear para prever estatísticas de desempenho de jogadores de futebol, fazendo várias perguntas ao “ChatGPT” como:

**Perguntas:**

* Há alguma forma de obter uma previsão mais realista alterando os valores de interpolação?
* Tendo em conta que se pretende, através de um conjunto de dados, prever um resultado futuro de um jogador, como é que obtemos uma previsão mais parecida com a realidade?
* Como diminuir a diferença entre as previsões feitas e os resultados reais?
* E em termos de precisão do resultado, como é que eu sei qual é o melhor valor para usar na interpolação?

O "ChatGPT" propôs uma solução que envolvia a combinação de dois conceitos diferentes do que estávamos a pensar usar nos primórdios do projeto, sendo estes a “Interpolação polinomial” e a “Validação Cruzada”.

"Validação Cruzada" é um método valioso de otimização não linear essencial por exemplo em "Machine Learning" que não foi ensinado na UC.

**Resposta:**

Para testar a precisão da interpolação, podes usar a validação cruzada. Este método consiste em dividir os valores em dois grupos, um de treino e outro de teste. Usa-se os dados de treino para criar uma função de interpolação e depois usamos o grupo de teste para avaliar a precisão da interpolação.

Segue um exemplo de como podes usar a validação cruzada para avaliar a precisão de diferentes números de interpolação.

# MATLAB code

x = [1 2 3 4 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29]; y = [2 1 1 1 2 3 2 1 3 2 2 2 1 1 1 1 3 2 4 2 1 1 1 1 1 0 1];

x\_new = 30;

% Divide se os dados num grupo de treino e teste idx = randperm(length(x)); train\_idx = idx(1:floor(length(idx)\*0.8)); % 80% para treino test\_idx = idx(length(train\_idx)+1:end); % 20% para teste

x\_train = x(train\_idx); y\_train = y(train\_idx); x\_test = x(test\_idx);

y\_test = y(test\_idx);

% Testar diferentes números para a interpolação num\_points\_to\_test = [2, 4, 8, 16];

rmse = zeros(length(num\_points\_to\_test), 1);

for i = 1:length(num\_points\_to\_test) % Interpolar usando os dados de treino

x\_interp = linspace(min(x\_train), max(x\_train), num\_points\_to\_test(i)); y\_interp = interp1(x\_train, y\_train, x\_interp, 'pchip');

% Avaliar usando os dados de teste

y\_pred = interp1(x\_interp, y\_interp, x\_test, 'pchip');

rmse(i) = sqrt(mean((y\_pred - y\_test).^2)); end

% Plotar os resultados

figure; plot(num\_points\_to\_test, rmse, 'o-'); xlabel('Number of interpolation points'); ylabel('RMSE');

title('Cross-validation results');

Este código usa 80% dos dados para treino e 20% para teste. Testa diferentes números de interpolação e calcula a raiz do erro quadrático da média entre os valores previstos e os valores atuais usados para teste. O RMSE é uma medida de quão os valores previstos estão longe dos valores atuais.

Plotando o RMSE para os diferentes números de interpolação, conseguimos ver qual é o ponto com o menor valor de erro no grupo de teste e, portanto, esse será o que terá as melhores previsões sobre os dados novos.

# -----------------------------------------------------------------------------------------------------------------

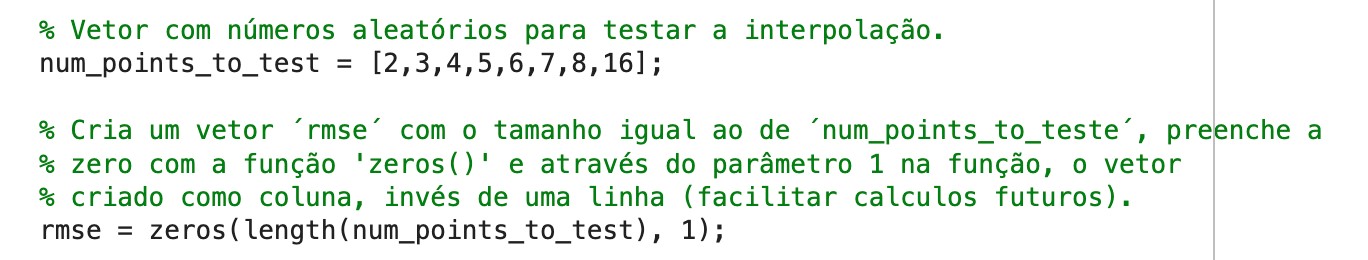
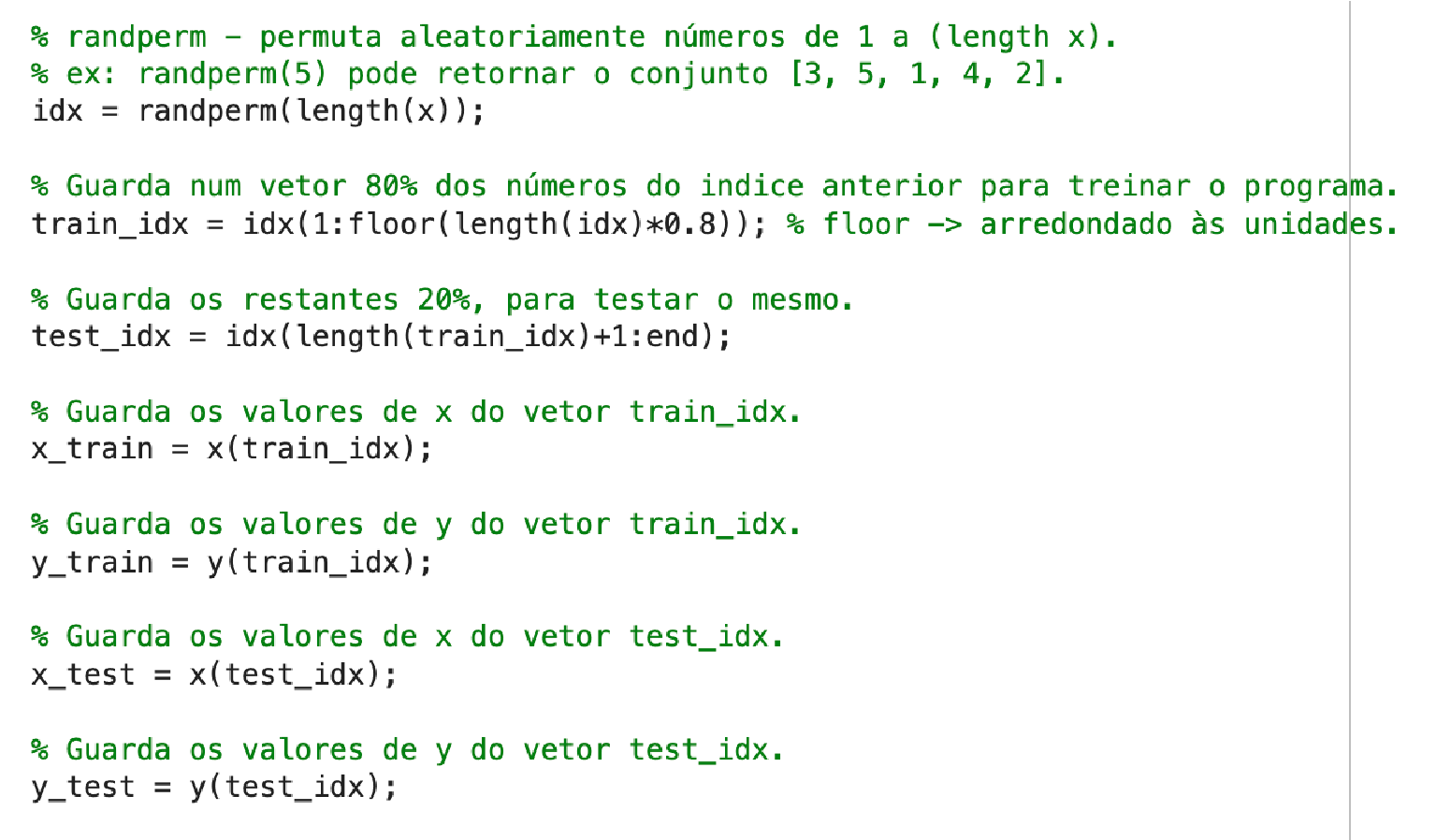
Animados com a possibilidade de finalmente obter uma previsão precisa, seguimos as orientações do "ChatGPT" e realizamos o modelo proposto.

# Codificação do Problema

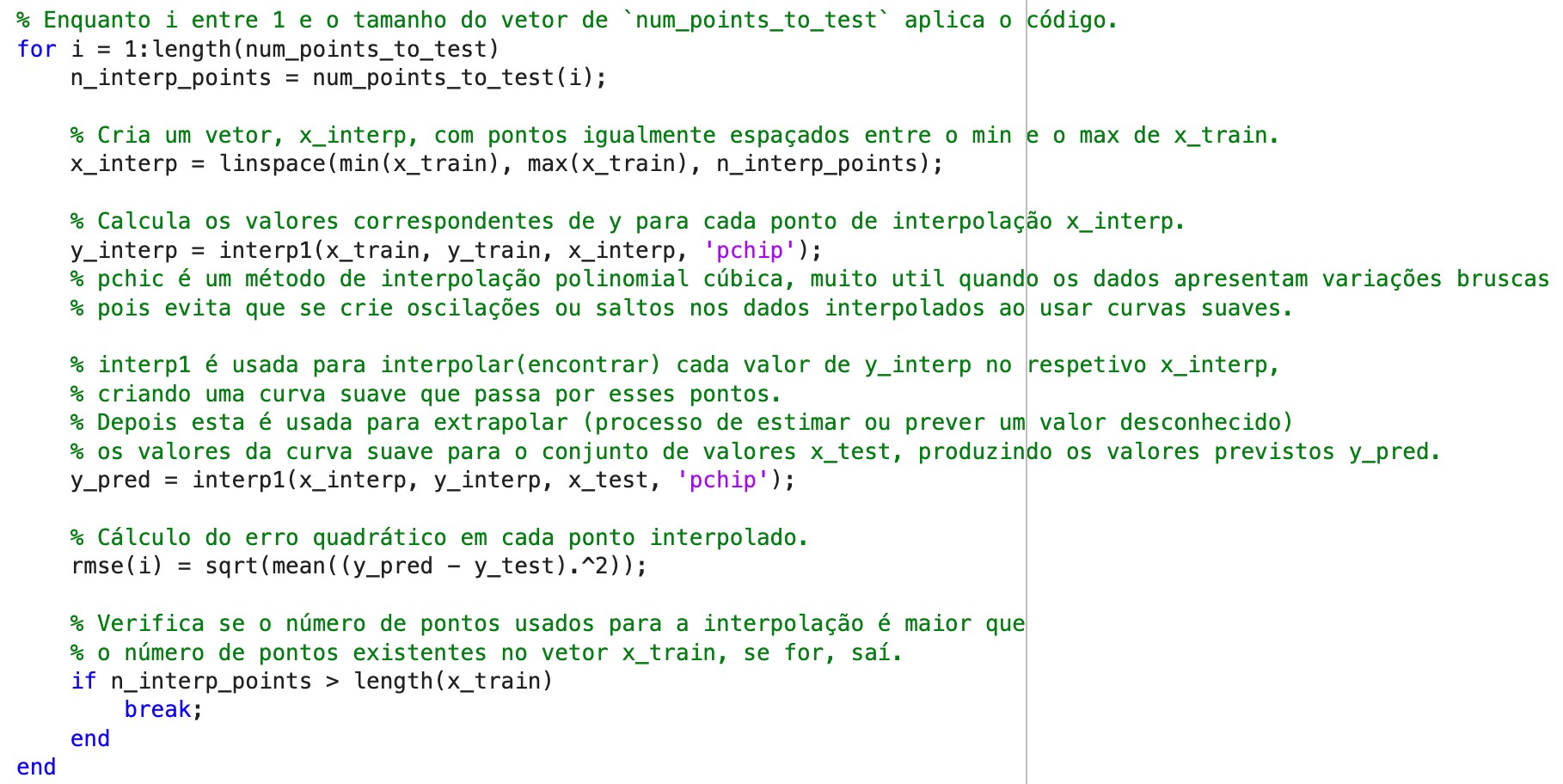
Uma vez escolhidos os subtemas, iniciamos a codificação em “Matlab”. Contudo, como parte do nosso modelo envolvia o uso de algumas rotinas não lecionadas, decidimos pedir ajuda ao “ChatGPT”.

De forma a trazer uma utilidade a este modelo, iremos prever quantos golos o Cristiano Ronaldo vai marcar no jogo 30 com o auxílio dos dados de uma temporada passada.

A estratégia proposta pela inteligência artificial é bastante simples. Primeiro, os dados são divididos em dois grupos: um grupo de treino, que corresponde a 80% dos dados, e um grupo de teste, composto pelos restantes 20%. O programa é treinado com o grupo de treino e, em seguida, é testado com o grupo de teste para prever resultados futuros com base no que foi aprendido durante o treino. Esta abordagem impede que o programa se adapte excessivamente aos dados usados durante o treino.

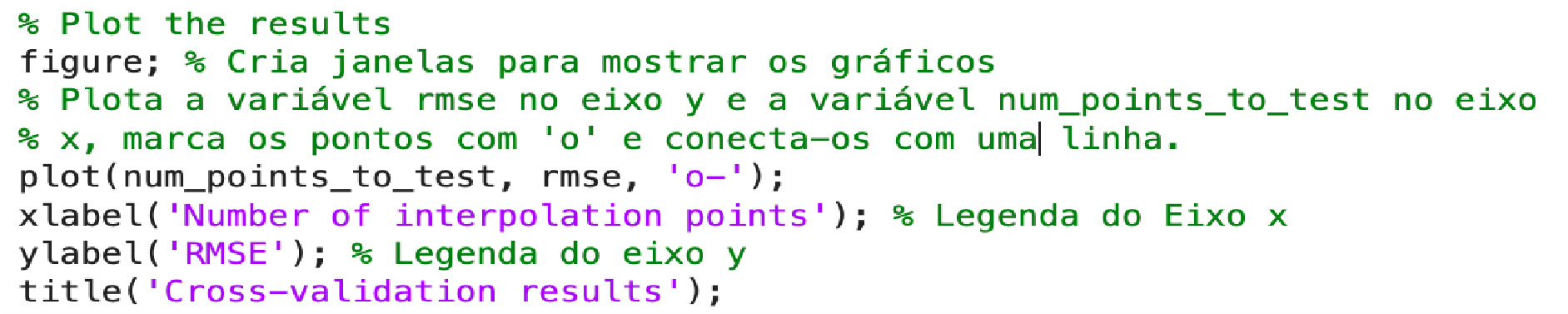


Posteriormente, sobre um conjunto de pontos escolhidos aleatórios e após a criação de um vetor ‘rmse’, o código executa um “loop” sobre o conjunto de pontos de treino para interpolar os dados usando um método de interpolação de polinomial do “Matlab”.



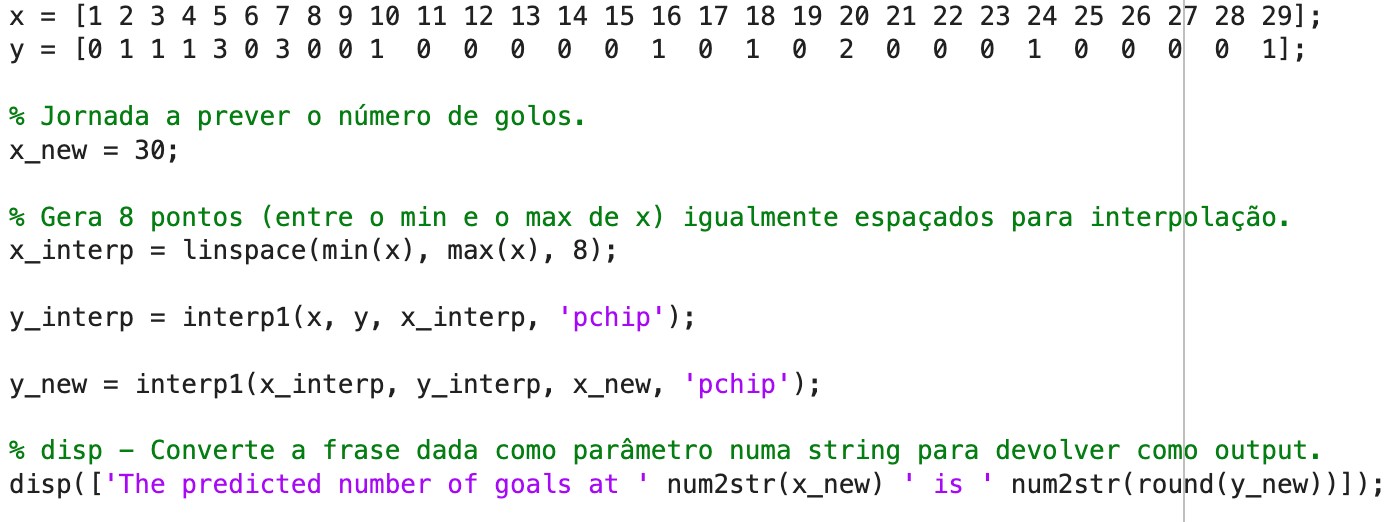
**Correção ao “ChatGPT”:** Inicialmente a proposta do “ChatGPT” não tinha em conta o número de pontos a interpolar e o tamanho do vetor “x\_train” por isso, adicionamos um loop “if “de forma que, o número de pontos para interpolar não ultrapassasse o tamanho do vetor evitando que sejam escolhidos mais pontos dos que há disponíveis.

O código, em seguida, traça o “EMQ” para uma gama de diferentes números de pontos de interpolação para avaliar a precisão da interpolação para diferentes pontos e é usado para determinar a quantidade ótima de pontos a serem usados para a interpolação.

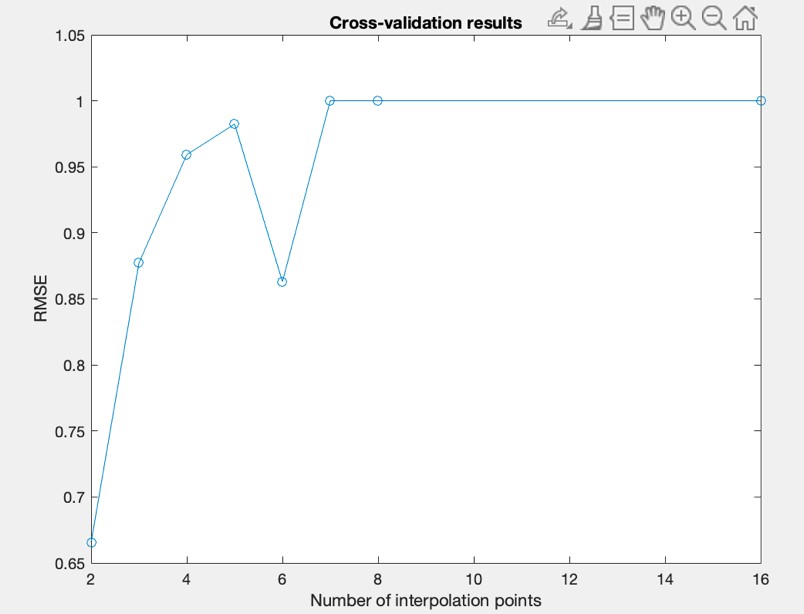
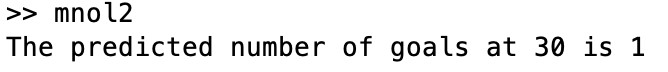


De modo a obter se resultados mais fidedignos, o código foi executado várias vezes, uma vez que, os dados obtidos para o grupo de treino e de teste iam variando, dado que eram aleatoriamente escolhidos afetando assim os resultados obtidos. Por isso, o valor que escolhido deve ser o que aparece mais vezes com um “EMQ” inferior.

Por fim, realizamos num script à parte com os mesmos dados, uma interpolação tendo em conta o número de pontos ótimos estimado anteriormente.



No caso exemplo, podemos ver que o modelo estimou que o Cristiano iria marcar 1 golo no jogo 30, que por usa vez, foi a situação que realmente aconteceu na temporada usada para estudo.



# Resultados variados

Para além do caso anterior, fizemos vários testes para diferentes estatísticas que um jogador obtém em uma partida. Assim, obtemos os seguintes dados:

**Diogo Leite Número de alívios no jogo 21.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 |  | 9 | | | 10 | |
| 5 | | 5 | | 0 | | 2 | | 0 | | 2 | | 4 | | 6 |  | 8 | | | 2 | |
|  | | | | | | | | | | | |  | |  |  |  | | |  | |
| 11 | | 12 | | 13 | | 14 | | 15 | | 16 | | 17 | | 18 |  | 19 | | | 20 | |
| 1 | | 5 | | 8 | | 5 | | 4 | | 3 | | 1 |  | 3 |  | 1 | |  | 6 | |
| Jogo 21: 3 alívios    Previsão do modelo para 5 interpolações: 6    Falha por 3        **Rafael Leão Número de golos no jogo 22.** | | | | | | | | | | | |  |  | |  | | |  |  | |
| 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 |  |  | 8 |  | | 9 | 10 | |
| 2 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 |  |  | 1 |  | | 0 | 1 | |
|  | | | | | | | | | | | |  |  | |  | | |  |  | |
| 11 | 12 | | 13 | | 14 | | 15 | | 16 | | 17 |  | 18 | | 19 | |  | 20 |  | 21 |
| 1 | 0 | | 0 | | 0 | | 0 | | 1 | | 0 |  | 2 | | 0 | |  | 1 |  | 0 |

Jogo 22: 0 golos

Previsão do modelo para 6 interpolações: 0

# Dificuldades encontradas

Inicialmente o “ChatGPT” propôs-nos uma solução muito pouco coerente tendo em conta o nosso caso de estudo e perdemos algum tempo para arranjar uma alternativa para que o resultado fosse de encontro com o que esperávamos. Para isso, questionamos o “ChatGPT” várias vezes, sobre novas alternativas e formas de otimizar o mesmo.

Uma outra dificuldade encontrada foi o facto de a proposta de trabalho ser totalmente revolucionária para todo o grupo, visto que, nenhum de nos tinha feito algum trabalho do género tornando se de todo um autêntico desafio.

# Conclusões

Após a análise dos resultados obtidos em casos de estudo, constatamos que o modelo utilizado nem sempre corresponde aos resultados esperados devido a diversos fatores, como a presença de "outliers" ou falta de informações sobre os jogos. Além disso, a escolha do número de pontos para interpolar influencia significativamente a precisão da previsão. Para superar essas dificuldades, é necessário selecionar casos de estudo que variem menos, cartões ou número de golos, e escolher jogadores com muitos registos de jogos para melhorar o modelo.