
a)


b)


c)


d)

a) Ana Catarina Canelas (a93872) b) Ana Filipa Pereira (a89589)
c) Carolina Santejo (a89500) d) Raquel Costa (a89464)

Ano Letivo 2020/2021

Mini-projeto Matlab 1 - Versão B

Grupo 19

Introdução

O tema escolhido pelo grupo foi a variação da Bitcoin em Euros ao longo de um intervalo de tempo. Após uma pesquisa de diversos temas no site https://www.statista.com/, escolhemos este devido à sua representação gráfica e comportamento oscilatório, além de ser um tema atual e relacionado com a área tecnológica. A amostra foi recolhida no site https://cointelegraph.com.br/bitcoin-price-index, inicialmente, o tamanho da amostra era 366 (dias, desde 9/11/19 até 9/11/20), mas depois o grupo decidiu reduzir para 153 (1 semestre), de modo a facilitar as análises e conceções de modelos. O objetivo é tentar encontrar o modelo que melhor se aproxime dos pontos da amostra e reflita o comportamento dos dados, seja x o número do dia e $f(x)$ o valor da Bitcoin em € nesse dia.

Modelo Polinomial

Como estamos a lidar com valores na ordem dos milhares decidimos fazer o escalonamento do problema, dividindo o $f(x)$ da amostra por mil, de forma a evitar a utilização de valores muito grandes e facilitando a análise e a interpretação do problema.

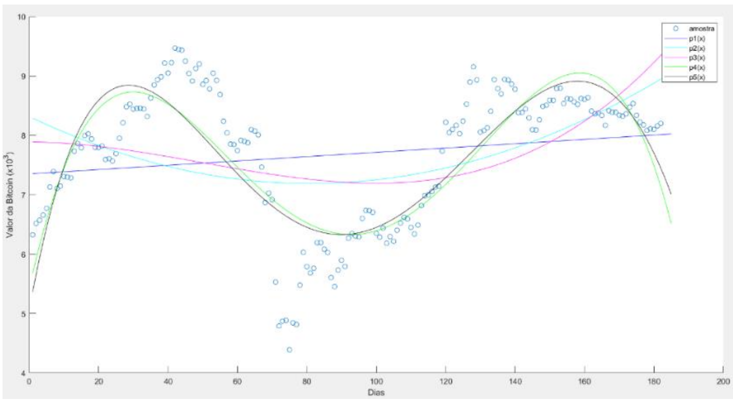


Figura 1- Modelos polinomiais de grau 1,2,3,4 e 5.

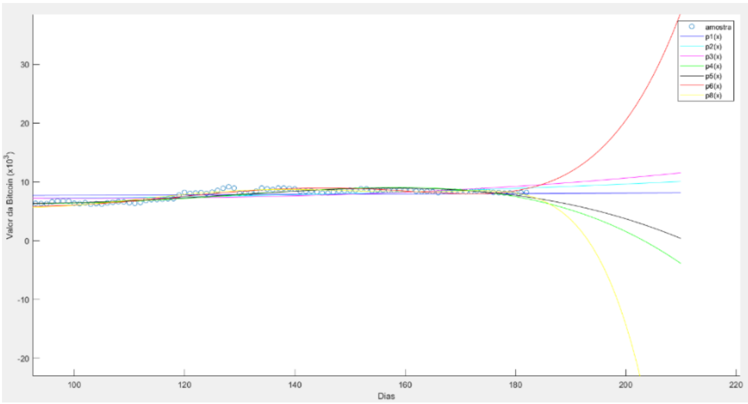


Figura 2- Aumento do intervalo e adição de novos polinómios

$p_1(x) = (0.0036 + 7.350.6x) * 10^3$	$S_1 = 229.8935 * 10^6$
$p_2(x) = (0.0002 - 0.0279x + 8.317\,x^2) * 10^3$	$S_2 = 197.0038 * 10^6$
$p_3(x) = (-0.0002x - 0.0001x^2 + 7.8883\,x^3) * 10^3$	$S_3 = 192.5301 * 10^6$
$p_4(x) = (0.0001x - 0.0066x^2 + 0.2633x^3 + 5.4187\,x^4) * 10^3$	$S_4 = 82.0355 * 10^6$
$p_5(x) = (0.1x - 0.0001x^2 - 0.0088x^3 + 0.3211x^4 + 5.0522\,x^5) * 10^3$	$S_5 = 80.1513 * 10^6$

Pela análise do gráfico (Figura 1) e dos erros de cada polinómio, percebemos que o polinómio que melhor aproxima a nossa amostra é o p_5 (o seu erro é o menor entre todos os obtidos). No entanto, o grupo, ao realizar testes computacionais reparou que todos os polinómios com grau superior ou igual 6, originavam o aviso "*Warning: Polynomial is badly conditioned*". Isto acontece, pois, apesar de polinómios com grau mais elevado possuírem um erro menor (passam por mais pontos da amostra) e, no gráfico até aparentarem um melhor ajuste, estes também são mais oscilatórios. Se verificarmos o comportamento do polinómio para mais pontos do que m (tamanho da nossa amostra), verificaremos que não é de todo desejável. Aliás, este tipo de modelos são conhecidos como ***overfitted models***, ou seja modelos que se ajustam bastante bem a um conjunto de dados mas revelam-se bastante ineficazes em determinar novos resultados. É possível observar oscilações mais acentuadas, no gráfico da Figura 2, no qual colocamos no eixo do x mais pontos do que ***m*** e adicionamos dois novos polinómios de grau 6 e 8 respetivamente.

Modelo Não Polinomial

Após várias tentativas do grupo de modo a encontrar um modelo $M(x)$ que melhor se aproxima aos pontos da amostra, escolhemos os seguintes com base no seu resíduo.

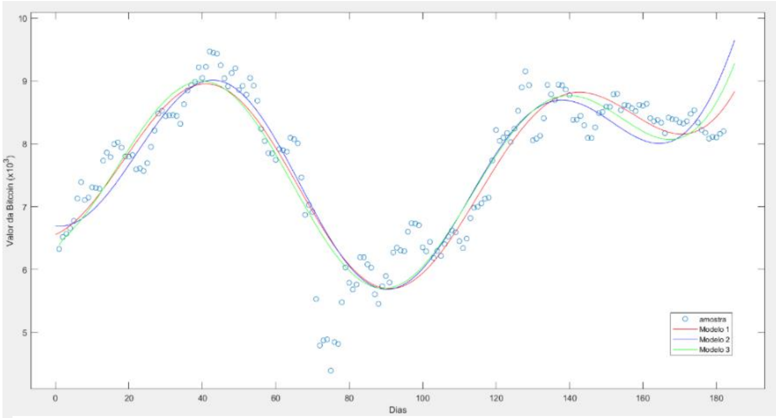


Figura 3 - Modelos Não Polinomiais

Uso de 3 Parâmetros: (Vetor inicial : c0 = [1 1 1])	
$M1(x) = c(1) * sin((x./15) + 5) + c(2) * (x.^2) + 10 * c(3) * cos(x./80)$	$S = 40.5486 * 10^6$
$M2(x) = c(1) * sin((x./14) + 1.5) + c(2) * (x.^2) + c(3) * cos(x./76)$	$S = 45.0386 * 10^6$
Uso de 4 Parâmetros: (Vetor inicial: c0 = [1 1 1 1])	
$M3(x) = c(1) * sin((x./15) + 5616) - c(2)./(x.^2) + c(3) * cos(x./200) + (c(4) * x.^5)$	$S = 40.3070 * 10^6$

Para construir os modelos, o grupo observou que o comportamento da amostra revelava oscilações do tipo sinusoidais. Portanto, decidimos usar as funções sin e cos como Φ_1, Φ_3, respetivamente. Além disso, como as oscilações não se mantêm constantes, mas parecem que vão "crescendo", adicionamos a função $x^2(\Phi_2)$. Inicialmente, usamos 3 parâmetros, mas depois, por curiosidade, utilizamos 4, para testar o seu resíduo e o seu ajuste aos pontos da amostra. Neste caso, para descobrir Φ_4, fomos incrementando o expoente de x e vimos que com o 5 era quando obtínhamos o menor resíduo. Para trabalhar com os sin e cos, foi preciso alguma pesquisa para compreender como ajustar estas funções. Ao longo do trabalho vimos que alterando algum aspeto de uma das funções implicava uma alteração na outra. Algo que é importante referir é que muitas vezes o grupo construía um modelo que aparentava ajustar-se muito bem a $f(x)$, mas depois o seu resíduo era grande. Concluímos que isto acontecia, devido aos ***outliers***, i.e há pontos que estão mais afastados e que fazem com que o resíduo aumente, portanto a estratégia do grupo foi escolher um modelo que se aproxime tanto destes pontos mais afastados como daqueles que estão mais concentrados numa zona, de forma a encontrar um equilíbrio e diminuir, consequentemente, o resíduo. Observando os resíduos de $M1(x)$ e $M3(x)$ vemos que ambos são modelos adequados.

Modelos Lineares – Redução da Amostra

De modo a explorar uma abordagem diferente do problema, decidimos considerar a média do valor da *Bitcoin* em Euros por semana ao longo de um ano ($f(x)$), em que o seu total são 52 semanas (x). Como estamos a fazer a média dos dias e, sendo assim, a excluir pontos, vemos que existem menos *outliers*. Será de prever que o resíduo seja menor e seja mais fácil de encontrar um modelo adequado.

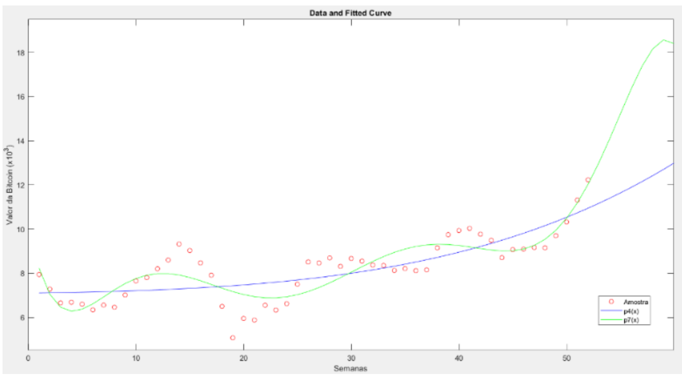


Figura 5 – Modelo Polinomial: Polinómio de grau 4 e 7

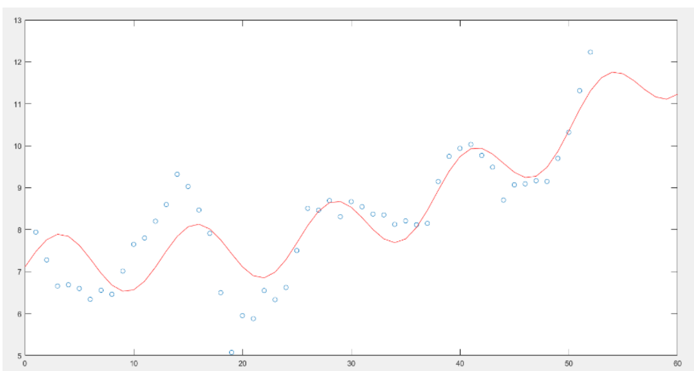
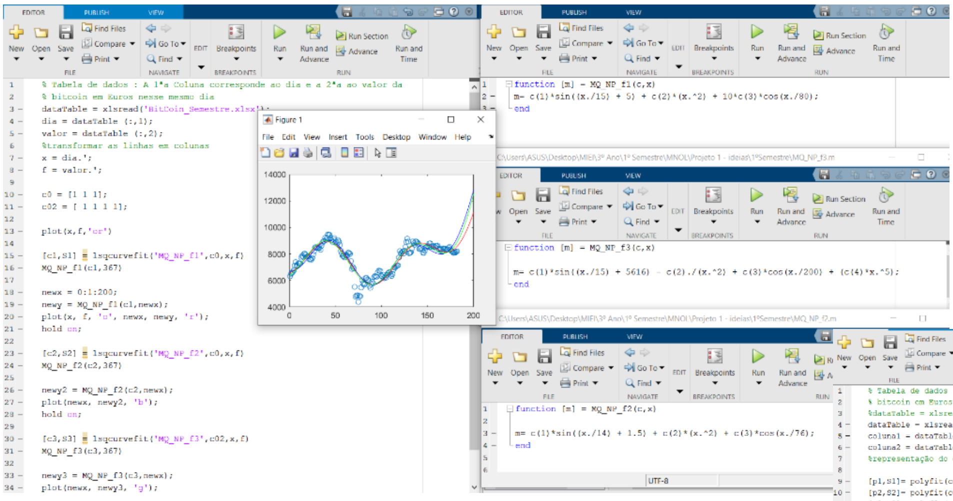


Figura 4 - Modelo Não Polinomial

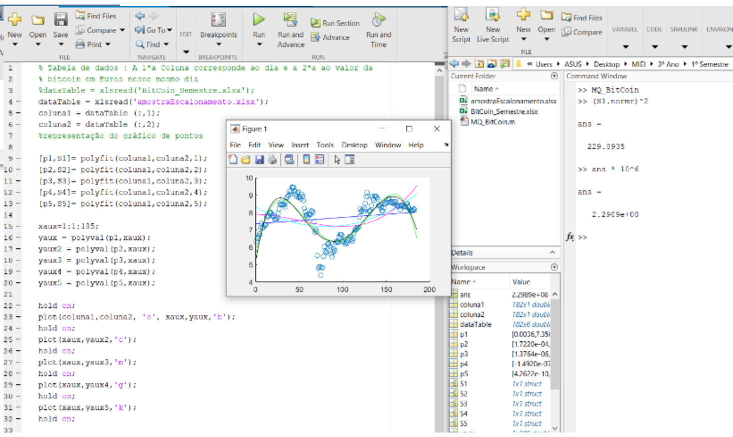
Agora com a amostra reduzida, vemos que os resíduos obtidos com os **polinómios $p_4(x)$ e $p_7(x)$** (Figura 5), são respetivamente $S_4 = 41.7631 * 10^6$ e $S_7 = 25.9097 * 10^6$. Comparativamente com os modelos polinomiais da primeira amostra utilizada, vemos que os resíduos são menores, comprovando assim o que tínhamos previsto, uma vez que temos menos pontos e não estão tão dispersos. O $p_7(x)$ apesar de parecer ajustar-se muito bem aos pontos, revela uma grande oscilação, logo concluímos que provavelmente não irá prever nem aproximar-se dos próximos pontos (podemos verificar isto aumentando a amostra).

O **Modelo Não Polinomial** (Figura 4) obtido foi: $M(x) = (c(1) * sin(x./2)) + 7 + c(2) * cos(x./8 - 0.8) + c(3) * 5 * x.^2;$, cujo resíduo é $S = 27.9513 * 10^6$. Como esta amostra mostra a variação da Bitcoin num ano, já conseguimos perceber melhor o “crescimento” de $f(x)$, reforçando assim a inclusão de x^2. Além disso, vimos que foi muito mais fácil encontrar um modelo apropriado comparativamente com a amostra anterior. Concluímos que este é o melhor modelo porque tem um resíduo relativamente baixo, ajusta-se bem à amostra e consegue prever razoavelmente o comportamento dos dados

Scripts e Comandos



Em ambos os scripts (polinomial e não polinomial) foi utilizado o comando $xlsread(\text{‘}nomeFicheiro.xlsx\text{’})$ para importar o ficheiro Excel das amostras. No final, para observar as representações gráficas, foi necessário o comando $plot$ e o $polyval$ para escolher quais os valores a exibir.



Modelo não polinomial

Para calcular o modelo não polinomial foi utilizado o seguinte comando:

$$[c, S] = lsqcurvefit('MQ\_NP\_f', c0, x, f)$$

Sendo $'MQ\_NP\_F'$ a função não polinomial e $c0$ o vetor dos parâmetros iniciais, ambos definidos anteriormente e (x, f) os pontos ao quais se deseja aproximar a função.

Após a sua execução é possível observar o seu resíduo S e os novos parâmetros em c.

Modelo polinomial

Para calcular o modelo polinomial foi utilizado o seguinte comando:

$[p, s] = polyfit(coluna1, coluna2, n)$

Sendo $(coluna1, coluna2)$ os pontos ao quais se deseja aproximar o polinómio p e n o seu grau. O resíduo pode ser calculado fazendo $(s.normr)^2$.

Conclusão

Em suma, foram apresentados os modelos mais adequados de amostras diferentes. Ainda fizemos o escalonamento do problema de modo a facilitar o tratamento de dados. Apesar de só termos apresentado 2 amostras, realizamos os mesmos testes computacionais para outras amostras de 365 dias e 12 meses. Ainda que os resultados obtidos tenham sido bastante satisfatórios, provavelmente ainda seria possível alcançar valores mais favoráveis. Tudo isto contribuiu para um melhor aproveitamento da UC porque permitiu-nos chegar a várias conclusões através da curiosidade aliada às experiências realizadas.