

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG - TPHCM
KHOA CÔNG NGHỆ THÔNG TIN



Đồ án 1: Nội suy

Phép nội suy đa thức

Học Phần: Phương pháp tính

Giảng viên hướng dẫn:

Thầy Nguyễn Đình Thúc

Thầy Trần Hà Sơn

Lớp: 23CLC02

Sinh viên thực hiện:

23127379 – Thái Minh Huy

Ngày 19 tháng 07 năm 2025

MỤC LỤC

1	Thông tin	4
1.1.	Thông tin cá nhân	4
1.2.	Thông tin cấu hình máy tính thực hiện	4
1.3.	Những thư viện cần thiết	4
1.4.	Thông tin đồ án	4
2	Giới thiệu	5
2.1.	Giới thiệu về nội suy đa thức Spline	5
2.2.	Những công cụ được sử dụng	6
2.2.1.	Công cụ sử dụng:	6
2.2.2.	Những module được sử dụng	6
3	Bài 1	9
3.1.	Đề bài	9
3.2.	Câu hỏi 1	10
3.3.	Câu hỏi 2	11
4	Bài 2	16
4.1.	Đề bài	16
4.2.	Thực hiện	16
4.2.1.	Lấy những ngày lớn hơn không	16
4.2.2.	Trải đều các ngày trong tuần	18
4.2.3.	Nhận xét	19
4.3.	Câu hỏi 1	20
4.4.	Câu hỏi 2	20
4.5.	Câu hỏi 3	22
4.6.	Câu hỏi 4	23
5	Tài liệu tham khảo	26

MỤC LỤC HÌNH ẢNH

Hình 1: Nhiệt dung của các dung dịch Axit Sulfuric	9
Hình 2: Phác thảo biểu đồ đường cong Spline với tập dữ liệu gốc	11
Hình 3: Hồi quy đa thức bậc một với tập dữ liệu gốc	12
Hình 4: Hồi quy đa thức bậc hai với tập dữ liệu gốc.....	12
Hình 5: Hồi quy đa thức bậc năm với tập dữ liệu gốc.....	13
Hình 6: Hồi quy đa thức bậc mười với tập dữ liệu gốc	13
Hình 7: Hồi quy đa thức và nội suy Spline bậc ba tự nhiên với tập dữ liệu gốc.....	14
Hình 8: Hồi quy đa thức khi tăng bậc càng lớn (cụ thể là bậc 40).....	15
Hình 9: Dữ liệu gốc về số ca nhiễm mới	17
Hình 10: Biểu đồ Covid-19 với dữ liệu gốc và Spline bậc ba tự nhiên.....	18
Hình 11: Dữ liệu gốc với Spline bậc ba tự nhiên nhưng trải đều số ca nhiễm trong tuần	18
Hình 12: Spline bậc ba tự nhiên trên dữ liệu Covid-19 nếu nội suy 50 điểm	21
Hình 13: Spline bậc ba khi nội suy quá nhiều nút.....	21
Hình 14: Spline nội suy trên toàn bộ dữ liệu.....	22
Hình 15: Spline bậc ba với SMA.....	23
Hình 16: Phương pháp Moving Average khi tăng kích thước cửa sổ	25

1 Thông tin

1.1. Thông tin cá nhân

MSSV	Họ tên	Email
23127379	Thái Minh Huy	tmhuy23@clc.fitus.edu.vn

1.2. Thông tin cấu hình máy tính thực hiện

- OS Name: Microsoft Windows 11 Home Single Language.
- Version: 10.0.26100 Build 26100.
- Processor: 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz, 2419 MHz, 4 core (s), 8 logical Processors(s).
- RAM: 8GB.

1.3. Những thư viện cần thiết

Trong đồ án này, em có sử dụng những thư viện, phiên bản như sau:

- Python: phiên bản 3.12.0.
- Anaconda: phiên bản 3.12.7. Base Kernel sử dụng là Python 3.12.7.
- Thư viện NumPy: dùng để xử lý tính toán khi sử dụng ma trận hoặc List của Python, phiên bản sử dụng: 1.26.4.
- Thư viện matplotlib: dùng để phác họa biểu đồ của dữ liệu gốc với dữ liệu khi sử dụng các phép toán nội suy (cụ thể là Spline bậc ba tự nhiên).
- Thư viện pandas: dùng để đọc dữ liệu trong file csv, phiên bản sử dụng là 2.2.2.
- Thư viện sklearn để dùng tính sai số tuyệt đối và sai số bình phương tối thiểu.

1.4. Thông tin đồ án

Ở phạm vi đồ án này, em sẽ được ứng dụng phần kiến thức toán xấp xỉ là nội suy đa thức Spline bậc ba để có thể phân tích biểu đồ dựa trên đường cong Spline và đánh giá, nhận xét. Sau đó sẽ trả lời các câu hỏi dựa trên phân tích đường cong của Spline và sai số của Spline so với dữ liệu thực tế.

2 Giới thiệu

2.1. Giới thiệu về nội suy đa thức Spline

Việc xây dựng một đa thức đi qua các điểm nội suy cho trước trong trường hợp n rất lớn là rất khó khăn và khó ứng dụng. Một trong những cách ta có thể khắc phục là trên từng đoạn liên tiếp của các cặp điểm nút nội suy ta nối chúng lại bởi các đường cong đơn giản như đoạn thẳng. Tuy nhiên, khi đó tại các điểm nút hàm sẽ mất tính khả vi. Do đó, người ta cố gắng xây dựng một đường cong bằng cách nối các đoạn cong nhỏ lại với nhau sao cho vẫn bảo toàn tính khả vi của hàm. Đường cong như vậy được gọi là đường Spline hay đường ghép trơn.

Các hàm trên các đoạn nhỏ thông thường là các đa thức và bậc cao nhất của các đa thức đó là bậc của Spline.

Định nghĩa 1:

Cho $f(x)$ xác định trên đoạn $[a, b]$ và một phép phân hoạch của nó: $a = x_0 < x_1 < x_2 = b$. Đặt $y_0 = f(x_0)$, $y_1 = f(x_1)$, $y_2 = f(x_2)$. Một Spline bậc ba nội suy hàm $f(x)$ trên $[a, b]$ là hàm $g(x)$ thỏa các điều kiện sau:

- $g(x)$ có đạo hàm cấp 2 liên tục trên $[a, b]$.
- $g(x) = \begin{cases} g_0(x), & x \in [x_0, x_1] \\ g_1(x), & x \in [x_1, x_2] \end{cases}$ ở đây $g_0(x), g_1(x)$ là các đa thức bậc ba.
- $g(x_0) = f(x_0) = y_0, g(x_1) = f(x_1) = y_1, g(x_2) = f(x_2) = y_2$.

Định nghĩa 2:

Cho $f(x)$ xác định trên đoạn $[a, b]$ và một phép phân hoạch của nó: $a = x_0 < x_1 < x_2 < \dots < x_n = b$. Đặt $y_k = f(x_k), k = 0 \dots n$. Một Spline bậc ba nội suy hàm $f(x)$ trên $[a, b]$ là hàm $g(x)$ thỏa các điều kiện sau:

- $g(x)$ có đạo hàm đến cấp 2 liên tục trên $[a, b]$.
- Trên mỗi đoạn $[x_k, x_{k+1}], k = 0 \dots n-1$, $g(x) = g_k(x)$ là một đa thức bậc ba.
- $g(x_k) = f(x_k) = y_k \forall k = 0 \dots n$.

Spline bậc ba tự nhiên

Điều kiện để xác định một Spline bậc ba tự nhiên là:

$$g''(a) = g''(b) = 0$$

$$g''(a) = g_0''(x_0) = 0$$

$$\Leftrightarrow 2c_0 + 6d_0(x_0 - x_0) = 0 \Rightarrow c_0 = 0$$

$$g''(b) = g_n''(x_n) = 0$$

$$\Leftrightarrow 2c_n + 6d_n(x_n - x_n) = 0 \Rightarrow c_n = 0$$

Giải hệ $AC = B$ tìm C với $C = (c_0, c_1, \dots, c_{n-1}, c_n)^T$ và

$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ h_0 & 2(h_0 + h_1) & h_1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2(h_{n-2} + h_{n-1}) & h_{n-1} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 3 \frac{(y_2 - y_1)}{h_1} - 3 \frac{y_1 - y_0}{h_0} \\ \dots \\ 3 \frac{y_n - y_{n-1}}{h_{n-1}} - 3 \frac{(y_{n-1} - y_{n-2})}{h_{n-2}} \\ 0 \end{bmatrix}$$

Sau khi tìm được $c_0, c_1, \dots, c_{n-1}, c_n$ thì các hệ số của $g_k(x)$ được xác định bởi:

$$\begin{cases} a_k = y_k \\ b_k = \frac{y_{k+1} - y_k}{h_k} - \frac{h_k}{3}(c_{k+1} + 2c_k), \forall k = 0 \dots n-1 \\ d_k = \frac{c_{k+1} - c_k}{3h_k} \end{cases}$$

2.2. Những công cụ được sử dụng

2.2.1. Công cụ sử dụng:

- Ngôn ngữ lập trình: Python.
- Text Editor: Visual Studio Code.
- Nền tảng: Jupyter Notebook.
- Quản lý và kiểm soát phiên bản mã nguồn (Source Control): Git và Github.
- Phương pháp lập trình: Lập trình hướng đối tượng (cho Spline bậc ba tự nhiên).

2.2.2. Những module được sử dụng

Trước khi thực hiện quá trình giải bài tập theo yêu cầu của đồ án, em đã đọc kỹ câu hỏi của hai bài tập được giao và điều yêu cầu sử dụng là nội suy Spline bậc ba tự nhiên cho nên em

đã tự xây dựng một lớp/module cho phép nội suy Spline, lớp NaturalCubicSpline sẽ có những phương thức như sau:

- *Constructor*: phương thức khởi tạo một object cho lớp NaturalCubicSpline, phương thức này sẽ nhận vào hai List:
 - X : mảng các giá trị của x .
 - Y : mảng các giá trị của y ứng với x .
 - Ngoài ra là thực hiện phương thức `compute_coefficients` để lấy hết các hệ số a, b, c, d của các đa thức nội suy Spline.
- *compute_coefficients*: phương thức này có nhiệm vụ thực hiện quá trình tính toán nhằm lấy hệ số a, b, c, d của các đa thức bậc ba khi thực hiện phép nội suy Spline.
 - Đầu tiên, chương trình sẽ thực hiện việc tính ma trận h có dạng là $1 \times (n - 1)$ với n là các cặp giá trị (x, y) từ hai bảng giá trị X, Y . Cụ thể, ứng với một phần tử của ma trận h sẽ được tính thông qua công thức truy hồi: trên đoạn $[x_k, x_{k+1}]$ thì $h_k = x_{k+1} - x_k \forall k = 0, \dots, n - 1$.
 - Bước tiếp theo là giải ma trận $alpha$, tương ứng là ma trận B theo định nghĩa, thực hiện vòng lặp chạy từ $[1, n - 1]$, phần tử đầu và phần tử cuối theo Spline bậc ba tự nhiên (c_0, c_n) sẽ có giá trị 0.
 - Sau khi có được ma trận $alpha$, tạo ma trận A có kích thước $n \times n$, khởi tạo toàn bộ giá trị cho các phần tử bằng 0, đồng thời cho phần tử đầu và cuối mang giá trị 1. Sau đó cũng chạy vòng lặp đi từ $[1, n - 1]$ và tính toán dựa trên các công thức truy hồi được đề cập ở phần định nghĩa. Thực hiện xong thì giải ma trận $A^{-1} \cdot B$ thì ta sẽ có được ma trận c , ma trận chứa danh sách là các hệ số c của các đa thức bậc ba của Spline.
 - Cuối cùng, ma trận a , chứa các hệ số a , sẽ lấy giá trị từ đầu đến cận cuối của danh sách Y . Chạy vòng lặp lần cuối cùng để lấy các hệ số b, d của đa thức.
- *interpolate*: phương thức này có nhiệm vụ tìm ra giá trị nội suy của x , nếu giá trị x cần được tính và $x \in [X[i]; X[i + 1]] \forall i \in [0, n - 1]$ thì giá trị y nội suy vào hàm đa thức bậc ba tương ứng và kết thúc hàm. Nếu duyệt hết mà x không thuộc bất kì khoảng nào thì phương thức sẽ trả về *None* (không mang giá trị gì).

Ngoài ra, để phục vụ cho bài 1 ý b , em có tự xây dựng thêm một module cho hồi quy đa thức (PolynomialRegression class), module này có những phương thức như sau:

- *Constructor*: phương thức khởi tạo một object cho việc xây dựng hồi quy đa thức và phương thức này cũng nhận vào hai list:
 - X : mảng các giá trị x .

- Y : mảng các giá trị của y ứng với x .
- *degree*: bậc của hồi quy đa thức.
- Phương thức này còn lưu lại hệ số của hồi quy đa thức thông qua phương thức *compute_coefficients*, giá trị lớn nhất và nhỏ nhất của mảng X .
- *compute_coefficients*: phương thức này có nhiệm vụ là tìm các hệ số của đa thức hồi quy.
 - Xây dựng ma trận A có kích thước $n \times (\text{degree} + 1)$.
 - Sau khi đã xây dựng ma trận A , tạo ra ma trận chuyển vị của ma trận A .
 - Lấy ma trận chuyển vị của A nhân với ma trận A và ma trận chuyển vị A nhân với ma trận Y .
 - Sau đó giải phương trình để tìm ma trận các hệ số của đa thức hồi quy.
- *predict*: phương thức này có nhiệm vụ nhận vào tham số x để dự đoán giá trị y .
- *min_max_normalize*: phương thức có nhiệm vụ chuẩn hóa hồi quy đa thức về giá trị đầu vào nằm trong đoạn $[0, 1]$, có nhiệm vụ xử lý dữ liệu đầu vào khi xây dựng mô hình hồi quy đa thức.
- *normalize_single*: phương thức có nhiệm vụ xử lý dữ liệu đầu vào để xây dựng mô hình hồi quy đa thức, đưa giá trị x nằm trong đoạn $[0, 1]$.

Ngoài ra, mỗi bài tập đều sẽ có một hàm helper nhằm hỗ trợ cho việc vẽ biểu đồ thông qua các phương thức từ thư viện matplotlib.

- Bài 1 hàm *plot_interpolation_results* sẽ biểu diễn trên biểu đồ là tập dữ liệu gốc về nồng độ và nhiệt dung riêng, đường cong Spline và đường cong hồi quy đa thức.
- Bài 2 hàm *plot_covid_interpolation* sẽ biểu diễn 4 phần, đầu tiên là tập dữ liệu gốc của Việt Nam, thứ hai là lọc dữ liệu bằng phương pháp lấy các ngày ghi nhận số ca nhiễm (lớn hơn 0) hoặc trải các số ca nhiễm trong tuần, thứ ba là đường cong Spline, cuối cùng là đường cong Moving Average hoặc LOESS (hồi quy cục bộ).

Để chạy chương trình, sau khi cài những thư viện cần thiết, chỉ cần nhấn *Run All* để chạy toàn bộ các ô dữ liệu nhằm thực thi chương trình và quan sát biểu đồ được vẽ cùng với tập dữ liệu gốc.

3 Bài 1

3.1. Đề bài

Nhiệt dung của các dung dịch Axit Sulfuric có liên quan đến nồng độ H_2SO_4 và tính chất này đã được đo lường cẩn thận trong phòng thí nghiệm với dữ liệu được trình bày trong hình vẽ dưới đây.

wt% H_2SO_4	C_p (kJ/kg·K)	wt% H_2SO_4	C_p (kJ/kg·K)
0.34	4.173	35.25	3.030
0.68	4.160	37.69	2.940
1.34	4.135	40.49	2.834
2.65	4.087	43.75	2.711
3.50	4.056	47.57	2.576
5.16	3.998	52.13	2.429
9.82	3.842	57.65	2.269
15.36	3.671	64.47	2.098
21.40	3.491	73.13	1.938
22.27	3.465	77.91	1.892
23.22	3.435	81.33	1.876
24.25	3.403	82.49	1.870
25.39	3.367	84.48	1.846
26.63	3.326	85.48	1.820
28.00	3.281	89.36	1.681
29.52	3.231	91.81	1.586
30.34	3.202	94.82	1.484
31.00	3.173	97.44	1.425
33.11	3.107	100.00	1.403

Hình 1: Nhiệt dung của các dung dịch Axit Sulfuric

Trong đó:

- wt% H_2SO_4 : phần trăm theo khối lượng tính theo công thức như sau:
 - $wt\% H_2SO_4 = \left(\frac{\text{khối lượng } H_2SO_4}{\text{tổng khối lượng dung dịch}} \right) \times 100\%$.

- C_p : nhiệt dung riêng ở điều kiện áp suất không đổi (constant pressure), công thức tính:
 - $q = m \cdot C_p \cdot \Delta T$.
 - Trong đó: q là nhiệt lượng, m là khối lượng, ΔT là độ biến thiên nhiệt độ.

Câu hỏi:

1. Áp dụng Spline bậc ba tự nhiên cho tập dữ liệu trên và trình bày biểu đồ đường cong Spline nội suy cùng với dữ liệu ban đầu.
2. Đưa ra nhận xét về khả năng khớp đường cong của Spline trong việc thực hiện điều mà hồi quy đa thức không thể làm được.

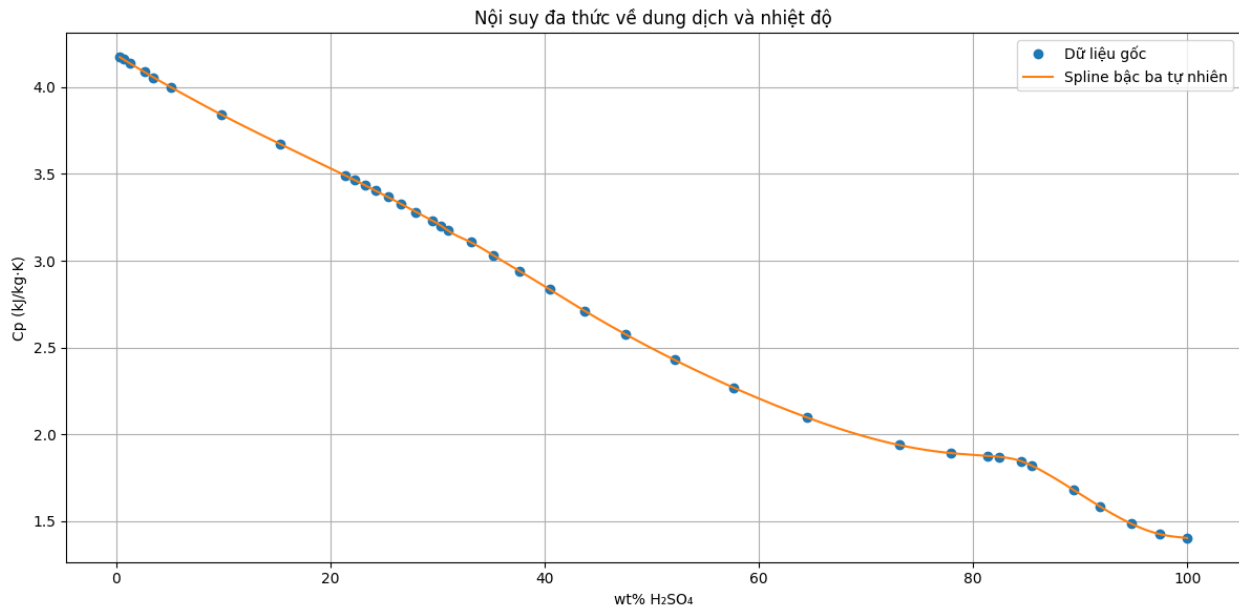
3.2. Câu hỏi 1

Trước khi thực hiện quá trình vẽ phác họa biểu đồ của đường cong Spline cùng với tập dữ liệu gốc được yêu cầu từ đề bài. Vì tập dữ liệu cho tập giá trị không nhiều, cụ thể là 38 cặp giá trị ($wt\%$, C_p) như đề bài cung cấp cho nên là em đã tự động nhập tay hết toàn bộ những dữ liệu của ($wt\%$, C_p). Những giá trị của $wt\%$ sẽ được bỏ vào danh sách *weight_percent* (phần trăm theo khối lượng), C_p sẽ được bỏ vào danh sách *heat_capacity* (nhiệt dung riêng ở điều kiện áp suất không đổi). Hai danh sách đó bắt buộc phải có cùng kích thước vì ứng với một giá trị $wt\%$ sẽ là một giá trị C_p (ánh xạ 1-1).

Sau khi đã bỏ hết các cặp giá trị ($wt\%$, C_p) tương ứng vào hai danh sách trên, em sẽ gọi một object tên *spline_model* để thực hiện công việc tính ra những đa thức bậc ba ở trên 38 giá trị của bảng dữ liệu gốc thông qua các phương thức em đã cài đặt để tính toán.

Sau khi tính toán, bây giờ là chương trình đã có 37 (tập dữ liệu gốc có 38 cặp giá trị cho nên sẽ có 37 đa thức) đa thức Spline bậc ba tự nhiên. Để có thể phác họa được lên đồ thị, trục hoành của đồ thị thì em cho là $wt\%$, trục tung là C_p , em chia thành 200 điểm để tính toán, thực hiện phép nội suy Spline bậc ba tự nhiên ở trên 200 điểm đó.

Khi này đây là em đã có tập dữ liệu $wt\%$ chia thành 200 điểm và ứng với đó là tập dữ liệu C_p sau khi đã nội suy Spline theo 200 điểm đó. Em đã phác thảo biểu đồ có đường cong Spline mà em đã xây dựng chung với tập dữ liệu gốc như trong hình ảnh dưới đây:



Hình 2: Phác thảo biểu đồ đường cong Spline với tập dữ liệu gốc

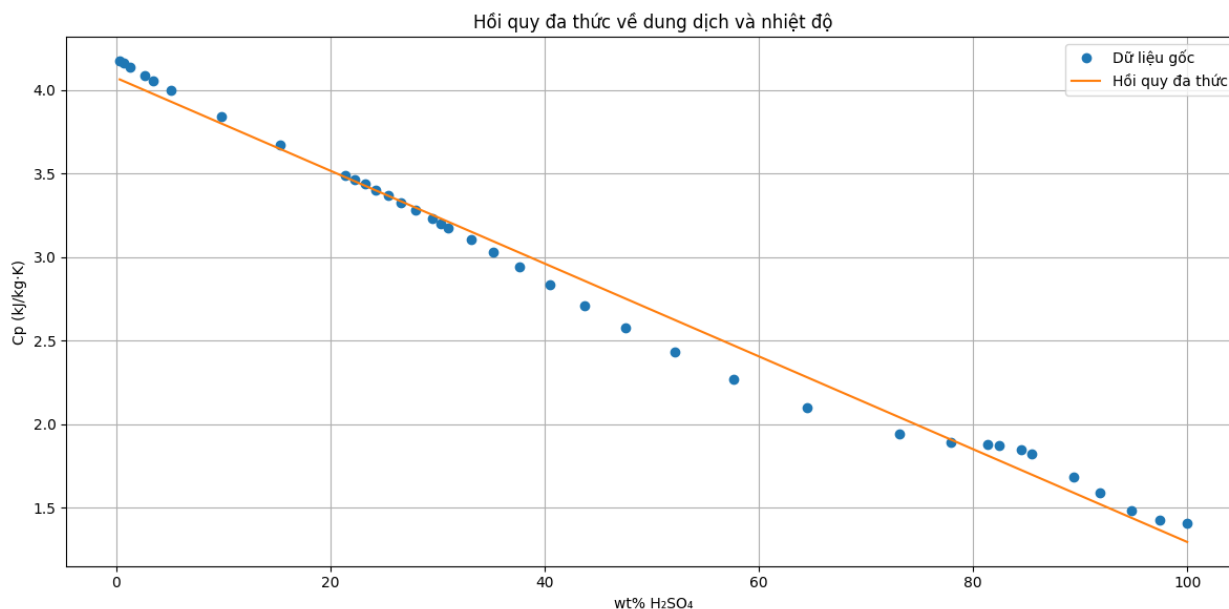
Nhận xét:

- Đường cong Spline sau khi phác thảo, cho ra kết quả là đường cong đi qua các điểm của tập dữ liệu gốc.
- Đường cong Spline có thể cho ra được kết quả trơn như trên là vì tại mỗi điểm $x_k, x_{k+1} \forall k = 0 \dots n - 1$, đa thức bậc ba là một hàm liên tục và khả vi cấp một và cấp hai, đã giúp cho đường cong mượt và không bị đứt đoạn tại các nút, khớp chính xác tại các điểm dữ liệu.
- Vì đường cong Spline chia nhỏ thành từng đoạn nhỏ cho nên Spline có thể khớp chính xác tại các điểm dữ liệu gốc nhằm hỗ trợ cho việc đánh giá tổng thể tốt hơn hay nói cách khác, đường cong Spline xử lý các điểm trên tập dữ liệu là xử lý cục bộ (local) hay vì toàn cục (global) cho nên Spline có thể tùy chỉnh để khớp dữ liệu.
- Đây là sai số mà chương trình tính toán được dựa trên dữ liệu thực tế với Spline bậc ba tự nhiên:
 - o Sai số tuyệt đối trung bình: $8.180590707764311e - 17 \sim 0.0000$
 - o Sai số bình phương tối thiểu: $2.0376210396350005e - 16 \sim 0.0000$
- Cả hai sai số đều cho thấy được rằng Spline khớp với tập dữ liệu gốc, dự đoán chính xác đường đi của tập dữ liệu gốc.

3.3. Câu hỏi 2

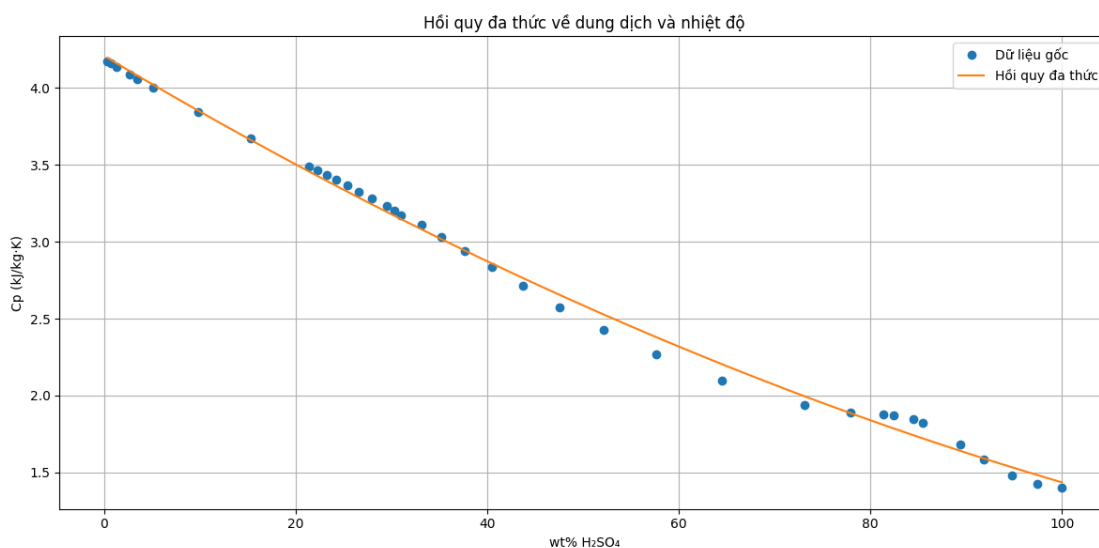
Để có thể đưa ra được nhận xét của cá nhân cho yêu cầu đề bài, em sẽ phác thảo đồ thị khi sử dụng mô hình hồi quy đa thức để có thể có cho mình một cái nhìn tổng quan nhất.

Em vẫn sẽ chia thành 200 điểm để quan sát, phác thảo biểu đồ khi sử dụng hồi quy đa thức cùng với tập dữ liệu gốc để quan sát. Dưới đây là hình ảnh khi phác khảo biểu đồ khi sử dụng mô hình hồi quy đa thức với tập dữ liệu gốc:



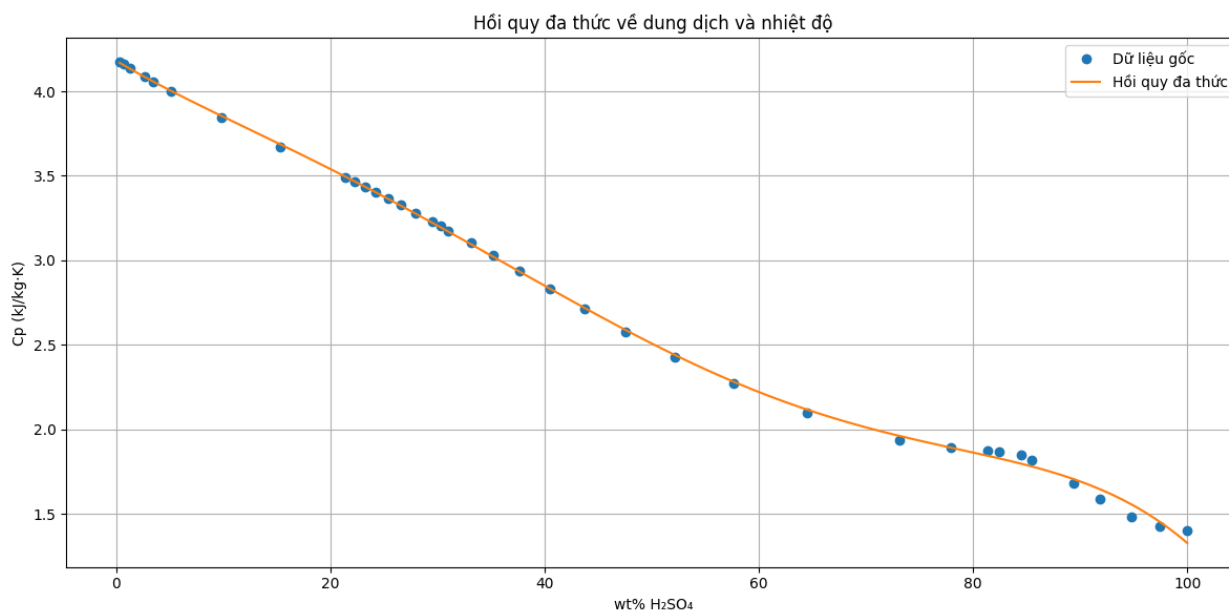
Hình 3: Hồi quy đa thức bậc một với tập dữ liệu gốc

Từ hình vẽ, ta có thể thấy được, khi sử dụng mô hình hồi quy đa thức có bậc là một thì chương trình sẽ vẽ ra cho ta một đường thẳng và đường thẳng này lại không có khớp hoàn toàn với dữ liệu gốc. Đường thẳng hồi quy chỉ đi qua rất ít điểm của tập dữ liệu gốc và trả ra kết quả không chính xác. Tuy nhiên nếu ta tăng bậc của mô hình hồi quy đa thức lên thì biểu đồ sẽ có những thay đổi như sau:

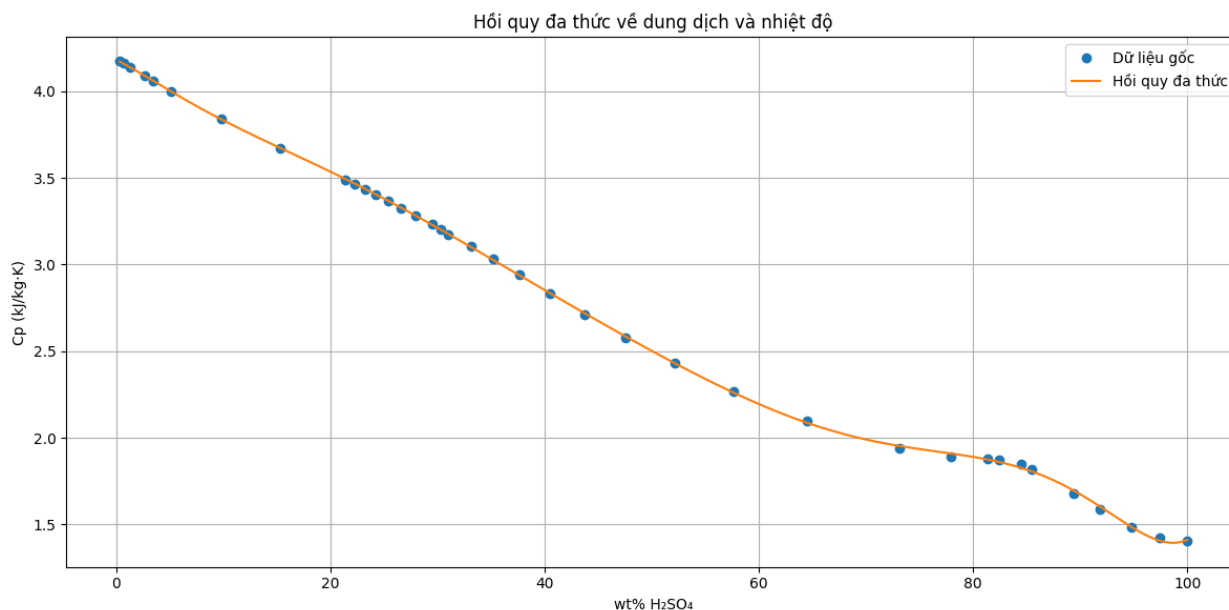


Hình 4: Hồi quy đa thức bậc hai với tập dữ liệu gốc

Khi ta tăng bậc của đa thức hồi quy lên bậc hai thì chương trình trả cho ta một đường cong và đường cong này bắt đầu đi qua những điểm của tập dữ liệu gốc. Tuy nhiên, đường cong này vẫn không hoàn toàn đi qua được những điểm khi wt% từ 20 trở đi cho đa thức này vẫn chưa hoàn toàn cho ta một kết quả tổng thể chính xác được để có thể rút ra kết luận.

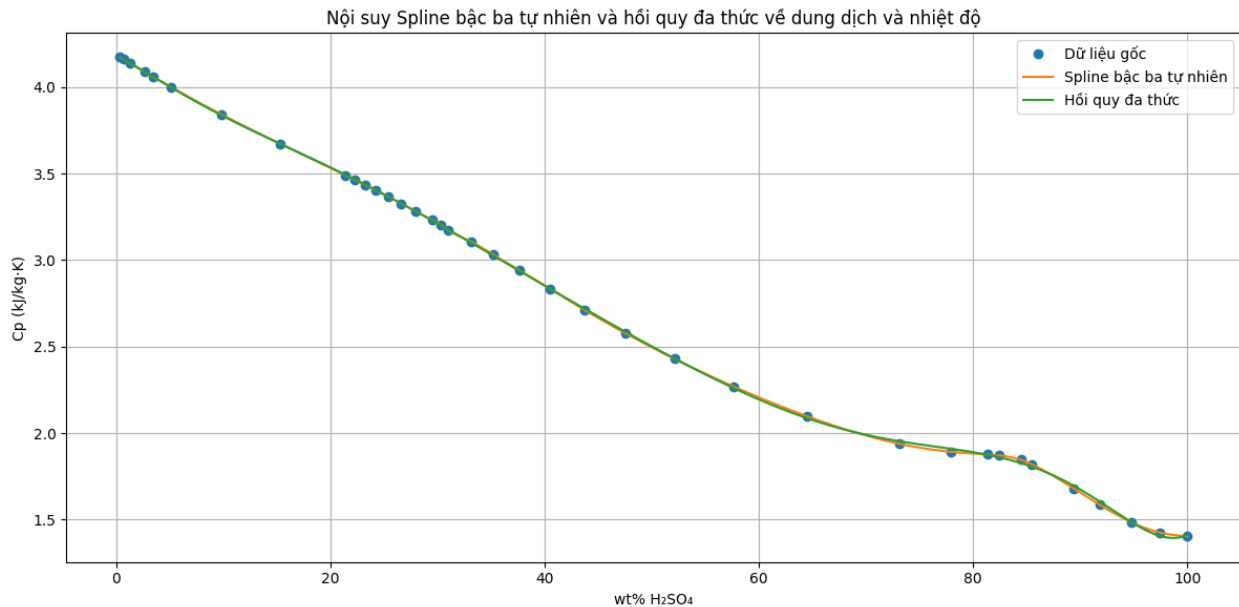


Hình 5: Hồi quy đa thức bậc năm với tập dữ liệu gốc



Hình 6: Hồi quy đa thức bậc mười với tập dữ liệu gốc

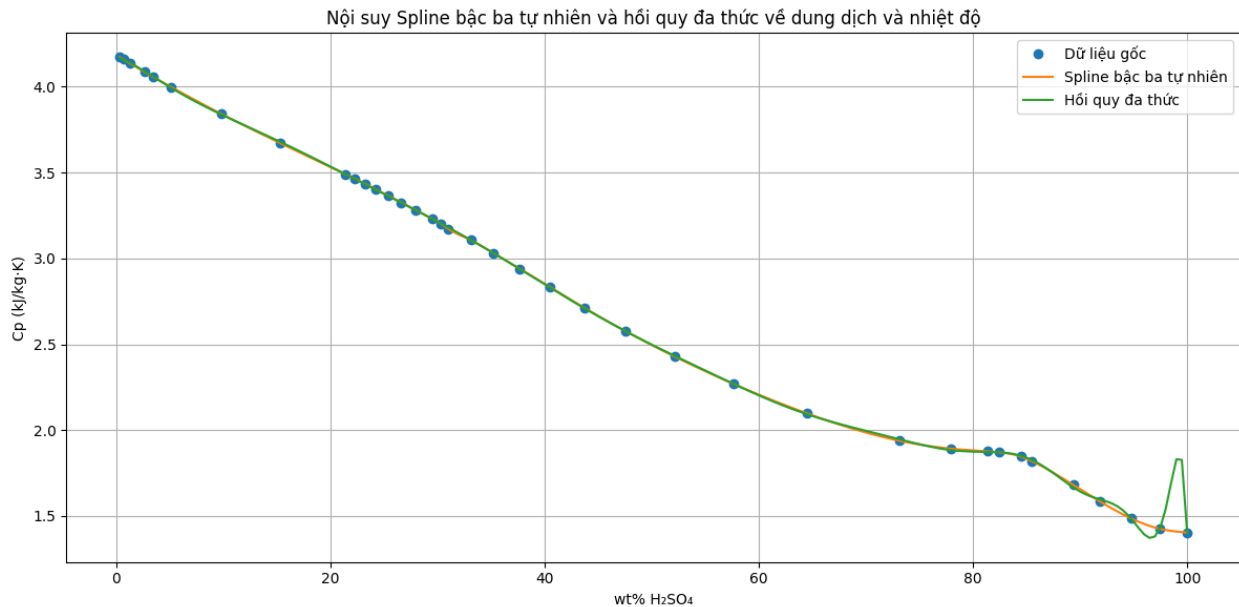
Khi ta tăng bậc của đa thức lên càng cao thì hồi quy đa thức đang vẽ ra đang dần đi qua hết tất cả các điểm của tập dữ liệu gốc để có thể đánh giá biểu đồ.



Hình 7: Hồi quy đa thức và nội suy Spline bậc ba tự nhiên với tập dữ liệu gốc

Nhận xét:

- Đối với nội suy Spline bậc ba tự nhiên, đường cong trả về trơn, liên tục hơn là bởi vì ứng với mỗi đoạn $[x_k, x_{k+1}] \forall k = 0 \dots n - 1$ là một đa thức bậc ba liên tục và có khả vi cấp một và cấp hai, liên tục về giá trị. Vì nội suy Spline chia các miền dữ liệu của tập dữ liệu gốc thành những đoạn nhỏ và khớp các điểm đó thông qua một đa thức bậc ba ở trên từng đoạn cho nên đường cong được trả về được trơn và liên tục. Nếu dữ liệu gốc có sự thay đổi xu hướng thì nội suy Spline có thể dễ dàng khớp dữ liệu bởi vì do nội suy Spline xử lý đường cong cục bộ nên sẽ ít gây ra những dao động mạnh về dữ liệu.
- Đối với hồi quy đa thức, ta càng tăng bậc của hồi quy đa thức cao thì sẽ giúp cho ta tạo ra một đường cong chính xác hơn. Nhưng khác với nội suy Spline khi mà Spline là xử lý cục bộ ở trên các nút nội suy thì hồi quy đa thức lại chỉ sử dụng một đa thức để có thể khớp hết toàn bộ dữ liệu. Việc chỉ sử dụng một đa thức duy nhất để khớp toàn bộ dữ liệu trên tập dữ liệu gốc sẽ cần bậc càng cao dẫn đến giá trị trả về rất lớn làm mất ổn định và dẫn đến trạng thái Overfitting (quá hợp). Ngoài ra, nếu như dữ liệu có những yếu tố ngoại lai hay có sự thay đổi xu hướng thì hồi quy đa thức sẽ không thể hiện chính xác biểu đồ so với Spline bậc ba tự nhiên.



Hình 8: Hồi quy đa thức khi tăng bậc càng lớn (cụ thể là bậc 40)

- Khi sử dụng đa thức hồi quy, ta càng tăng bậc của đa thức thì đa thức ấy sẽ khớp với tập dữ liệu gốc. Tuy nhiên điểm bất lợi của hồi quy đa thức là càng tăng bậc của đa thức thì số lượng phép toán sẽ rất lớn mà hồi quy đa thức lại sử dụng duy nhất một đa thức để điều khiển toàn bộ và đi qua hầu hết các điểm ở trên tập dữ liệu gốc thành ra phép toán sẽ trở nên cực kì lớn ở trong đa thức dẫn đến việc hồi quy đa thức sẽ bị dao động mạnh ở biên cuối và sẽ khiến cho việc dẫn đến trạng thái Overfitting trên tập dữ liệu. Đây chính là điểm đặc biệt khi em so sánh với phương pháp Spline bậc ba tự nhiên bởi vì Spline bậc ba chỉ xử lý ở cục bộ ở trên những đoạn $[x_k, x_{k+1}] \forall k = 0 \dots n - 1$. Spline không sử dụng một đa thức để khớp với tập dữ liệu gốc mà sử dụng những đa thức khác nhau ở trên các đoạn nhỏ và sẽ giúp cho đường cong trơn, mượt, liên tục và khớp ở trên tập dữ liệu gốc, đồng thời những tính toán của Spline sẽ không quá phức tạp hơn so với hồi quy đa thức. Những đa thức nhỏ của Spline sẽ xử lý nội bộ và khi những điểm đó có xu hướng không đồng đều thì Spline có thể uốn cong nhằm khớp với tập dữ liệu gốc.
- Dưới đây là sai số của hồi quy đa thức với bậc là mười so với Spline bậc ba tự nhiên:
 - Sai số tuyệt đối trung bình: 0.005943602436891597
 - Sai số bình phương tối thiểu: 0.008481567941456124
- Vì hồi quy đa thức chỉ sử dụng một hàm để điều khiển cho nên là sai số của phương pháp này sẽ lớn hơn nhiều so với Spline, sử dụng những đa thức bậc hai và bậc ba để nội suy.

4 Bài 2

4.1. Đề bài

Phân tích xu hướng dữ liệu các ca nhiễm của Covid-19 bằng Spline bậc ba.

Mục tiêu:

- Hiểu và áp dụng phương pháp Spline bậc ba nhằm làm trơn dữ liệu thời gian.
- Phân tích xu hướng và diễn biến của đại dịch Covid-19 dựa trên dữ liệu tần suất theo thời gian.

Dữ liệu:

- Số ca nhiễm hằng ngày (không tích lũy) của Covid-19.
- Khu vực: Việt Nam trong giai đoạn 2020-2021.

Câu hỏi:

- Đường cong Spline bậc ba có phản ánh đúng xu hướng tổng thể của dữ liệu không?
- Việc lựa chọn số lượng và vị trí các nút Spline ảnh hưởng như thế nào đến kết quả?
- Làm thế nào Spline giúp phát hiện các mốc dịch quan trọng mà dữ liệu thô có thể che khuất?
- Nếu so sánh với các phương pháp làm trơn khác (*moving average*, *LOESS*), có những ưu/nhược điểm gì?

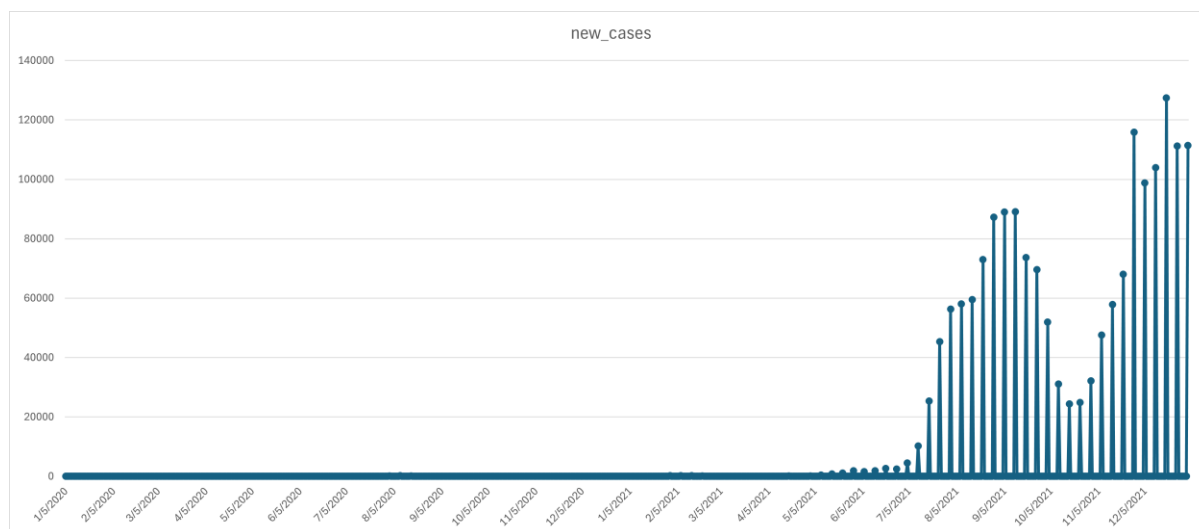
4.2. Thực hiện

Để có thể trả lời những câu hỏi theo yêu cầu, em sẽ thực hiện đọc file csv bằng cách sử dụng module pandas (file *data.csv*), chỉ lấy những giá trị thuộc khu vực Việt Nam (Vietnam), lấy những giá trị ngày (từ ngày 01-01-2020 đến ngày 31/12/2021 + 02/01/2022) cùng với đó là số ca nhiễm Covid-19 trong ngày (*new_cases*). Lọc bỏ đi những ngày mà thiếu dữ liệu hoặc những dòng có những giá trị không hợp lệ (NaN). Sau đó em có dùng hai phương pháp như sau để nội suy Spline.

4.2.1. Lấy những ngày lớn hơn không

Từ việc phân tích biểu đồ của Việt Nam trong thời kỳ dịch Covid-19 trong giai đoạn 2020-2021, em đã để ý một điều là dữ liệu đó chỉ cập nhật vào tất cả những ngày chủ nhật (cuối tuần). Những ngày còn lại không phải là không có dữ liệu mà là dữ liệu đó không được cập

nhật ngay lập tức mà chỉ được mặc định là đánh số 0. Số 0 ở những cột dữ liệu được đánh mang giá trị như vậy không có nghĩa là những ngày đó không có những ca nhiễm mới mà số ca nhiễm mới lại được tổng hợp vào những ngày chủ nhật. Mục đích của người làm ra dữ liệu này làm em không rõ tại sao lại làm như thế hay có những ẩn ý gì sâu xa nhưng đó là những gì mà em phân tích được sau khi em đã đọc và vẽ biểu đồ ở trên file Excel.

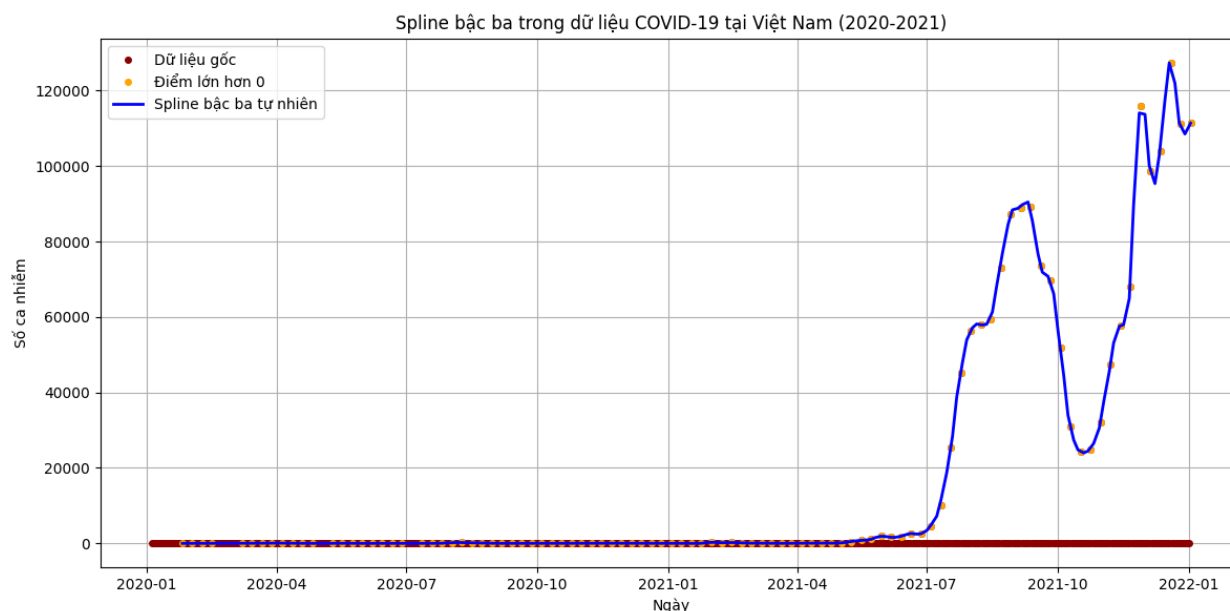


Hình 9: Dữ liệu gốc về số ca nhiễm mới

Sau khi đã đọc được thông tin của dữ liệu gốc, em chỉ lấy những cột dữ liệu mà số ca nhiễm lớn hơn 0 để bỏ vào hai danh sách. Danh sách đầu tiên là những ngày ghi nhận số ca nhiễm mới, danh sách thứ hai là ứng với ngày ghi nhận số ca mới thì sẽ là số ca đã được ghi nhận. Đây sẽ là hai dữ liệu được yêu cầu để phác họa biểu đồ và áp dụng nội suy để đánh giá. Sau đó, xây dựng các đa thức nội suy trên tập dữ liệu để chuẩn bị cho việc thực hiện tính toán. Em vẫn sẽ nội suy trên 200 điểm để phác họa biểu đồ.

Biểu đồ dưới đây sẽ phác họa ba phần:

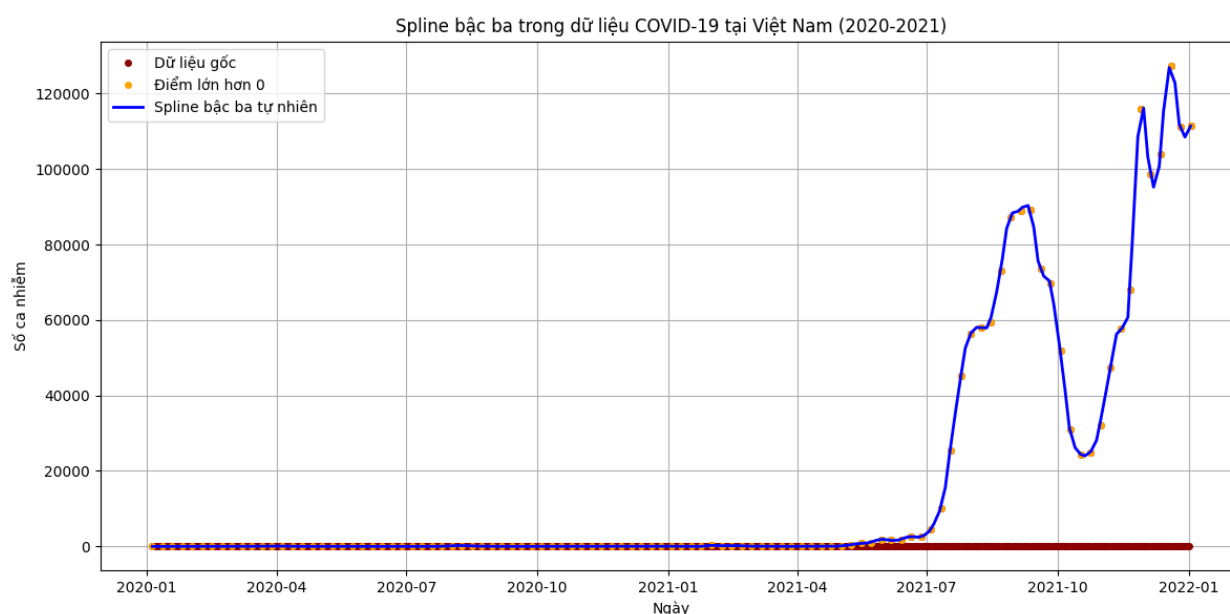
- Phần được biểu diễn bằng chấm màu đỏ sẽ là dữ liệu gốc của Việt Nam, dữ liệu này sẽ bao gồm những ngày mà không ghi nhận những số ca trong ngày và ghi nhận số ca vào những ngày chủ nhật.
- Phần được biểu diễn bằng chấm màu cam sẽ là dữ liệu của Việt Nam nhưng chỉ ghi nhận những miền dữ liệu mà số ca nhiễm mới là lớn hơn 0, những miền đó là từ những ngày chủ nhật.
- Phần được biểu diễn bằng đường cong màu xanh dương là đường cong Spline nội suy ở trên tập dữ liệu mà số ca nhiễm lớn hơn không.



Hình 10: Biểu đồ Covid-19 với dữ liệu gốc và Spline bậc ba tự nhiên

4.2.2. Trải đều các ngày trong tuần

Phương án thứ hai mà em đã sử dụng là em sẽ lấy dữ liệu cuối tuần được ghi nhận vào ngày chủ nhật đó, em sẽ trải đều rời rạc vào những ngày còn lại trong tuần đó. Vẫn lấy dữ liệu gốc sau khi tách từ file csv để lấy dữ liệu từ Việt Nam về, nội suy 200 điểm ở trên tập dữ liệu ngày và số ca nhiễm để xây dựng đa thức Spline bậc ba tự nhiên ở trên đây. Dưới đây là hình ảnh biểu đồ được phác họa:



Hình 11: Dữ liệu gốc với Spline bậc ba tự nhiên nhưng trải đều số ca nhiễm trong tuần

Biểu đồ trên cũng được biểu diễn bằng ba phần:

- Phần được biểu diễn bằng chấm đỏ là tập dữ liệu gốc về số ca nhiễm mới của Việt Nam, bao gồm những ngày được đánh dấu là số 0.
- Phần được biểu diễn bằng chấm cam là những điểm số ca nhiễm được trải đều vào những ngày trong tuần đó.
- Phần màu xanh là đường cong Spline được nội suy trên những điểm được biểu diễn bằng chấm cam.

4.2.3. Nhận xét

- Đường cong Spline bậc ba thông qua việc nội suy tại những điểm có số ca nhiễm mới lớn hơn không hay trải đều số ca nhiễm trong tuần đều cho ra một đường cong khớp với tập dữ liệu gốc, các đường cong trên đã khớp với những giai đoạn từ tháng 1 năm 2020 đến tháng 5 năm 2021 là những ngày mà nước ta chống dịch rất tốt khi mà ghi nhận số ca nhiễm rất ít so với các nước lân cận và đồng thời ở thời điểm, các ca nhiễm của nước ta khi phát hiện đều được nhanh chóng cách ly kịp thời nhằm tránh lây lan thành một ổ dịch bùng phát. Có thể nói đây là thời điểm mà nước ta rất tự hào khi đã chống dịch thành công, số ca tử vong được ghi nhận trong giai đoạn là không có. Ngoài ra là giai đoạn từ tháng 5-2021 đến tháng 12-2021, giai đoạn bùng phát những ổ dịch lớn, đặc biệt tại các đô thị trung tâm của đất nước.
- Tuy nhiên, đường cong đã bắt đầu có dấu hiệu dốc lên từ tháng 4-2021 đến tháng 7-2021, cụ thể là ngày dịp lễ 30/04-01/05/2021 là thời điểm người dân được nghỉ lễ và lên kế hoạch những ngày lễ đó về quê. Ở giai đoạn này, số ca nhiễm bắt đầu tăng lên và tăng đến chóng mặt, số ca nhiễm được ghi nhận quá nhiều dẫn đến hiện tượng các bệnh viện đã không thể tiếp nhận thêm nhiều bệnh nhân. Giai đoạn này cũng là lúc mà chỉ thị mọi người dân phải ở trong nhà được đưa ra nhằm để hạn chế gây ra những ca nhiễm mới, đây cũng là giai đoạn Việt Nam đã ghi nhận được những ca tử vong đầu tiên kể từ dịch Covid-19 được bùng phát.
- Dù giai đoạn trên rất khó khăn, nước ta với những chính sách chống dịch hợp lý đã giúp cho người dân có thể quay trở lại với nhịp sống bình thường, các mô hình giải trí và kinh doanh đều đã hoạt động trở lại từ tháng 10, 11 năm 2021.
- Điểm dịch đến tháng 11-2021 là những điểm dịch được xảy ra tại các tỉnh miền Nam, đỉnh dịch cuối cùng được thể hiện ở trên biểu đồ cũng là cao nhất được bùng phát tại các tỉnh miền Bắc.

4.3. Câu hỏi 1

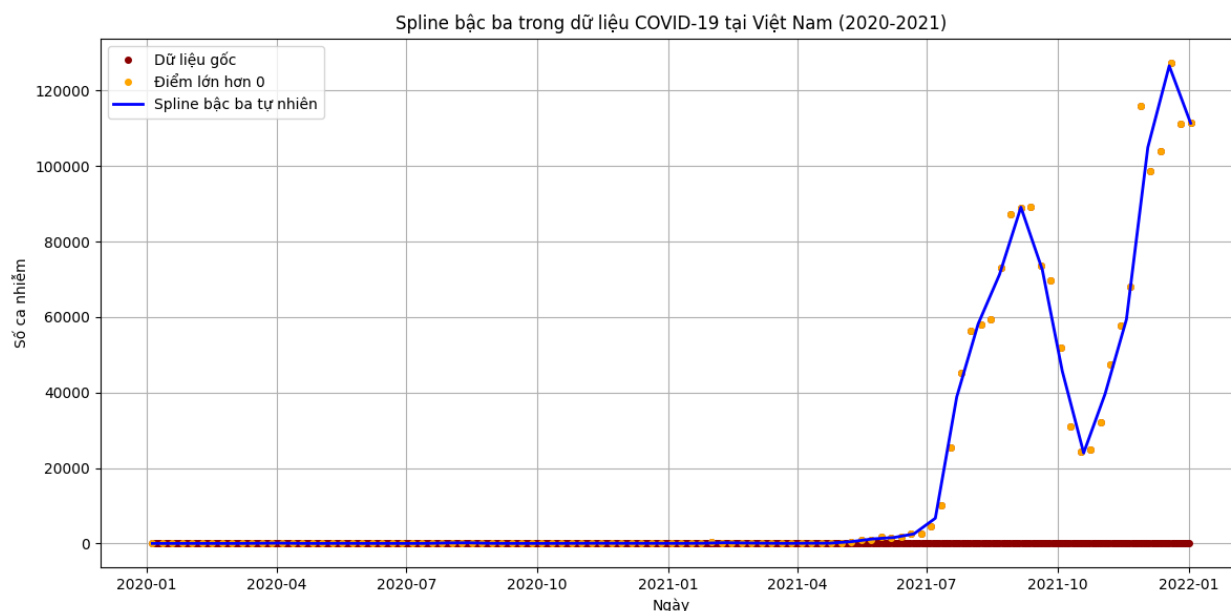
Dựa trên biểu đồ từ hai phương pháp nội suy em đã trình bày ở trên, hai biểu đồ cho đường cong Spline đều khớp với tập dữ liệu gốc về những số ca nhiễm mới trong ngày của Việt Nam. Đường cong Spline đã phản ánh đúng ba giai đoạn của Việt Nam trong thời kì chống dịch những năm đó, cụ thể như sau:

- Từ tháng 1/2020 đến tháng 5/2020, nước ta đã thực hiện việc cách ly những nạn nhân bị F0 ngay lập tức để tránh có thêm những ca nhiễm mới, thời điểm này nhà nước thực hiện việc cách ly xã hội cực kì chặt chẽ cho nên số ca nhiễm mới dường như không có quá nhiều sự dao động, thay đổi bất thường. Đây cũng là thời điểm mà nước ta tự hào khi là đất nước chống dịch rất tốt.
- Từ tháng 5/2021-9/2021 thì số ca nhiễm bắt đầu tăng lên nhanh chóng và là thời điểm mà đất nước phải thi hành nhiều chính sách chống dịch mới, nhiều lực lượng hỗ trợ nhằm giảm bớt những ổ dịch đang bùng phát tại các trung tâm đô thị, thành phố lớn (cụ thể là thành phố Hồ Chí Minh). Thông qua những chính sách được thực thi có kế hoạch cho nên từ tháng 10/2021, số ca nhiễm đã có xu hướng được giảm xuống và người dân tại thời điểm đó đã dần dần quay trở lại nhịp sống bình thường.
- Giai đoạn cuối cùng là bùng phát tại những thành phố tại tỉnh miền Bắc và số ca nhiễm mới lần này tăng lên và tăng lên cao nhất so với hồi tháng 9/2021.
- Thông qua nhận xét và đánh giá biểu đồ, đường cong Spline cho ta thấy được những giai đoạn nào mà nước ta ghi nhận số ca nhiễm ít, ghi nhận những số ca nhiễm nhiều. Đường cong Spline giúp ta đánh giá được tổng thể những giai đoạn mà số ca nhiễm mới được bùng phát chính xác tại các điểm được nội suy.

4.4. Câu hỏi 2

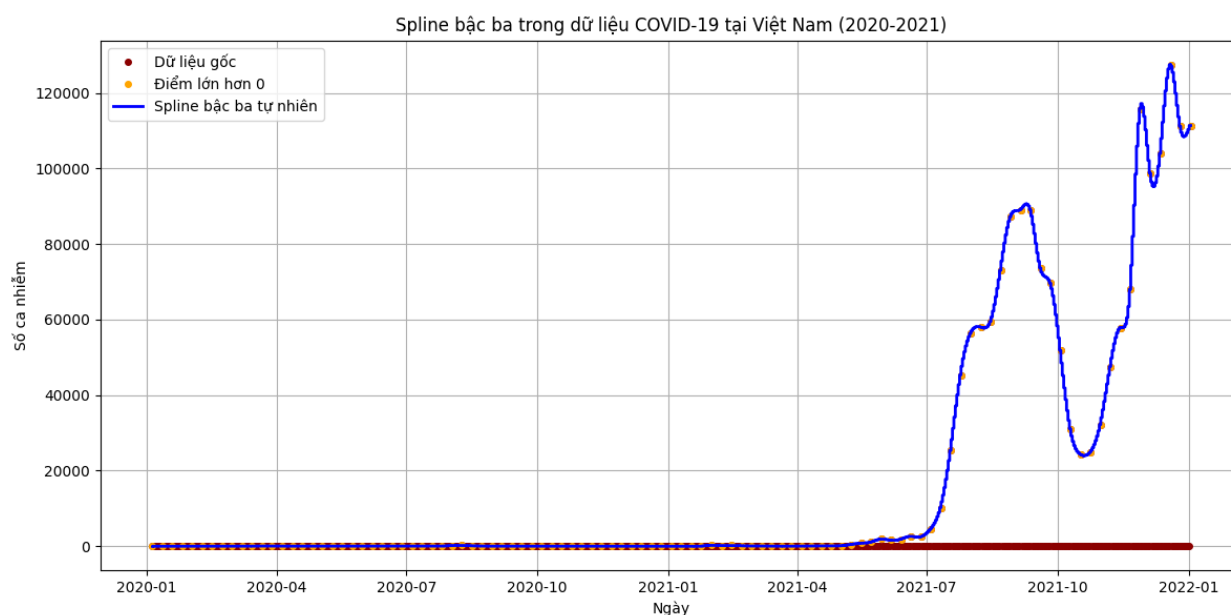
Việc lựa chọn số lượng nút Spline có thể sẽ ảnh hưởng đến kết quả như sau:

- Số lượng nút ít: nếu như mà ta nội suy Spline với số lượng nút quá ít, đường cong Spline sẽ không thể đánh giá đúng được kết quả tổng thể của dữ liệu và có thể đưa ra kết quả không thể chính xác, việc quá ít nút có thể sẽ khiến cho việc ta sẽ mất đi những chi tiết của dữ liệu như đỉnh dịch xảy ra ở giai đoạn nào. Dưới đây là hình ảnh minh họa cho thấy được nếu như ta nội suy quá ít điểm thì đường cong Spline sẽ có hình dạng như hình dưới đây, đường cong dù có thể đánh giá được số ca nhiễm đạt đỉnh ở đâu tuy nhiên kết quả sẽ không được chính xác vì đường cong sẽ bỏ qua và làm mất đi những đỉnh dịch bùng phát khác. Việc tính toán sẽ nhanh hơn do sẽ không phải nội suy nhiều nút nhưng kết quả để phân tích thì chính xác sẽ không thể cao so với trường hợp ta nội suy nhiều nút.



Hình 12: Spline bậc ba tự nhiên trên dữ liệu Covid-19 nếu nội suy 50 điểm

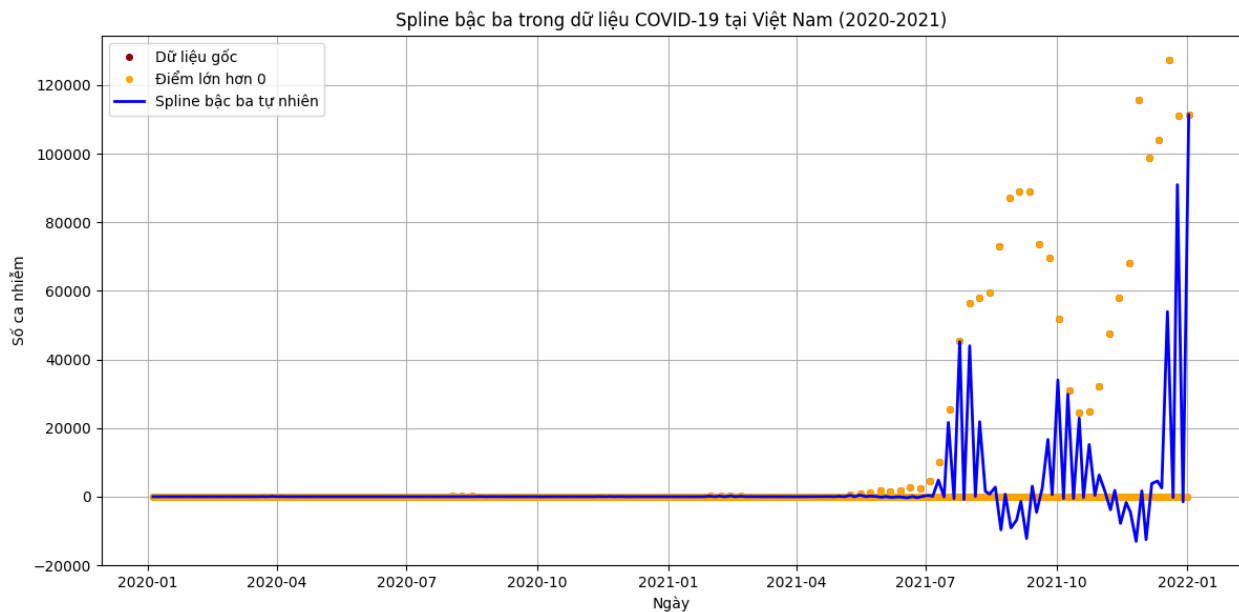
- Số lượng nút Spline nhiều có thể giúp cho ta có thể bám sát được dữ liệu thực tế hơn vì đường cong Spline sẽ nội suy trên nhiều điểm và có thể đánh giá được kết quả tốt hơn tuy nhiên đường cong sẽ cho cảm giác bị nhiễu, không trơn, liên tục và có thể cũng dẫn đến trường hợp Overfitting như hình dưới đây:



Hình 13: Spline bậc ba khi nội suy quá nhiều nút

Trong trường hợp nếu như ta lựa chọn nội suy ở trên toàn bộ tập dữ liệu, cụ thể nếu như mà lấy luôn cả những ngày bằng 0 thì đường cong Spline sẽ bị nhiễu và cho ra hình dạng

đường cong không thích hợp trong thực tế, ví dụ như hình ảnh dưới đây là nội suy trên toàn bộ tập dữ liệu. Đường cong Spline lần này nội suy trên toàn bộ điểm thành ra dẫn đến đường cong sẽ không phản ánh đúng được kết quả tổng thể của số ca nhiễm mới đã diễn ra, chưa kể đường cong lại có thể đi xuống ở mức âm dẫn đến trả về kết quả không thực tế vì số ca nhiễm không thể nào âm nên việc chọn vị trí các nút Spline có thể ảnh hưởng đến biểu đồ dẫn đến đánh giá tổng thể sẽ không còn chuẩn nữa.



Hình 14: Spline nội suy trên toàn bộ dữ liệu

Tổng kết lại: việc chọn số lượng các nút ít sẽ không thể đánh giá chính xác dữ liệu và có thể bị mất, quá nhiều sẽ sát với dữ liệu nhưng sẽ tạo cảm giác không còn trơn và liên tục được. Vị trí các nút Spline được chọn cũng ảnh hưởng vì nếu như ta chọn sai thì có thể trả về biểu đồ kết quả không chính xác thậm chí dẫn đến không hợp thực tế. Chính vì thế cho nên việc lựa chọn số nút nội suy cần phải được chú ý là nên chọn cân bằng, cụ thể là từ 200 – 300 nút để có trả ra biểu đồ trơn, mượt và liên tục đồng thời nắm bắt được xu hướng của dữ liệu.

4.5. Câu hỏi 3

Từ dữ liệu thô, ta có thể nhìn thấy một vấn đề mà em đã nói từ lúc đầu, có những ngày trừ ngày chủ nhật lại được đánh dấu là số 0. Số 0 trong trường hợp này có nghĩa là không phải những ngày đó không có những ca nhiễm bệnh mới mà là những ngày đó, người làm ra dữ liệu này lại không ghi nhận trong ngày đó mà họ đã tổng hợp những số ca nhiễm bệnh đó vào đúng duy nhất một ngày và đó là ngày chủ nhật. Dù đã tổng hợp số ca nhiễm vào một ngày duy nhất tuy nhiên nếu mà mình nhìn tổng thể ở trên tập dữ liệu đó thì ta rất khó xác

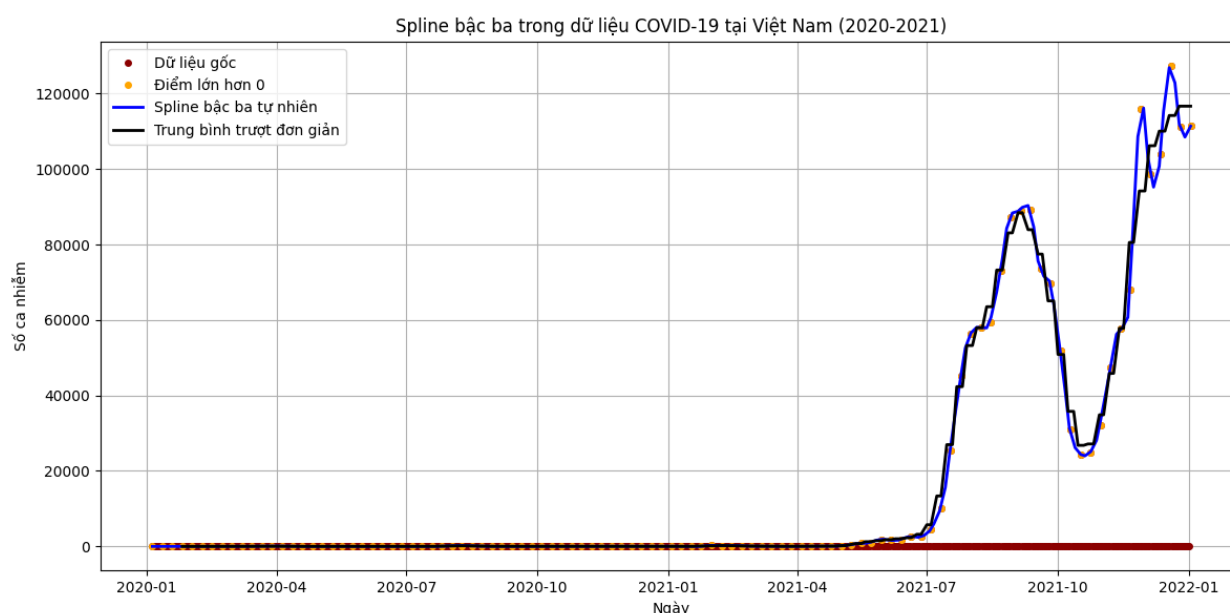
định được số ca nhiễm trong tuần đó hay về sau có xu hướng như thế nào, ta không rõ là số ca nhiễm có xu hướng tăng lên hay có xu hướng giảm đi. Đường cong Spline với tính chất là trơn, liên tục khả vi cấp một và cấp hai có thể giúp cho chúng ta kết nối những điểm ghi nhận số ngày mắc đó, nối lại với nhau bằng các đa thức trong cục bộ, kết hợp lại với nhau, bỏ qua những dao động nhỏ và tái dựng lại cho biết hiện tại xu hướng đã và đang như thế nào thông qua những điểm uốn cho biết là số lượng ca mắc đang tăng, điểm cực đại là số lượng ca mắc đạt lớn nhất và điểm cực tiểu cho ta biết là số ca mắc đang giảm xuống.

Ngoài ra, Spline còn giúp cho ta thấy và phát hiện điểm những mốc dịch quan trọng như từ giai đoạn tháng 5-9/2021 thì là điểm bùng phát cho biết số ca nhiễm bắt đầu tăng rõ rệt, đỉnh dịch là cho ta biết được số ca mắc đạt cực đại, suy giảm là cho ta biết được là số ca mắc bệnh đang giảm xuống. Cuối cùng là giai đoạn ổn định là cho biết là dịch bệnh đang ổn định, không có nhiều biến động.

4.6. Câu hỏi 4

Tại nội dung này, em chọn phương pháp làm trơn khác là Moving Average (Trung bình trượt) với biến thể là Simple Moving Average (Trung bình trượt đơn giản hay SMA) để so sánh với phương pháp xấp xỉ Spline đang được sử dụng trong đồ án.

Dưới đây là hình ảnh sử dụng phương pháp Moving Average:



Hình 15: Spline bậc ba với SMA

Định nghĩa:

Trung bình trượt đơn giản được tính bằng cách tính tổng giá trị trong một khoảng thời gian và chia cho số quan sát trong khoảng thời gian đó.

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n}$$

Trong đó:

- A là các giá trị dữ liệu.
- n là khoảng các giá trị dữ liệu hay còn được gọi là khung cửa sổ trượt.

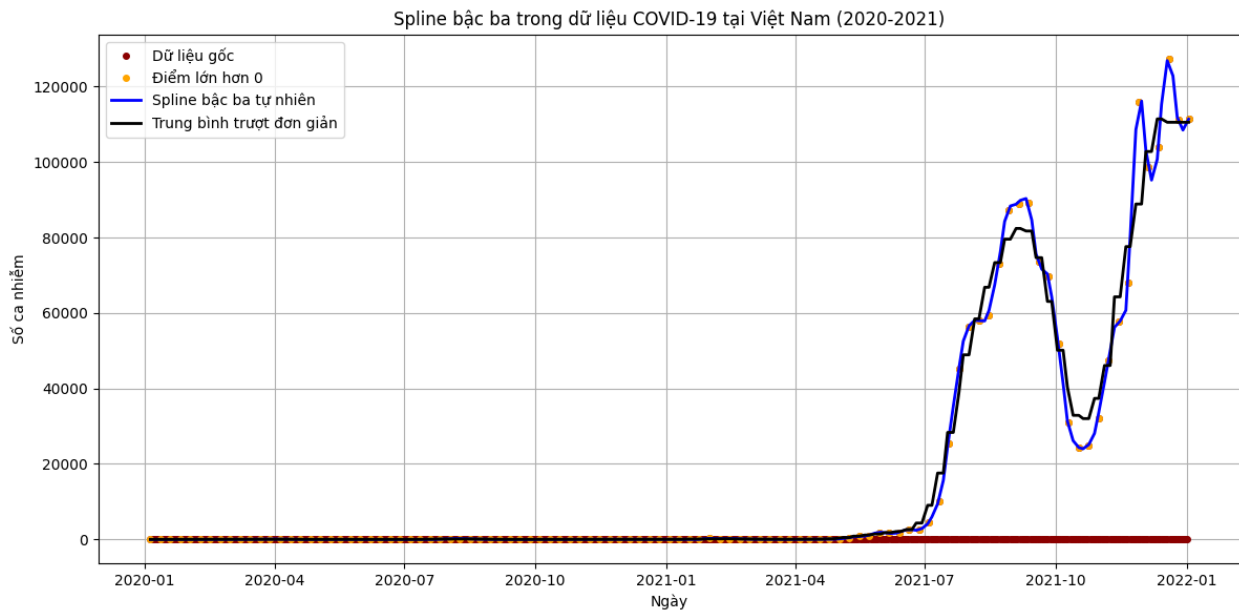
Ưu điểm của phương pháp Moving Average so với Spline:

- Moving Average sẽ tính trung bình dựa trên các điểm k gần nhất, nghĩa là khi dữ liệu được thêm vào ở biên cuối, cho dù những dữ liệu trước có thời gian rất lâu như 10 năm hay chỉ mới 1 tháng trước thì phương pháp này chỉ tính các giá trị trung bình dựa trên k dữ liệu trước đó (dựa trên quy tắc cửa sổ trượt – Sliding Window).
- Khác với phương pháp Spline bậc ba đang được sử dụng, phương pháp Moving Average với biến thể là Simple Moving Average là thuật toán dễ cài đặt bởi vì chúng ta chỉ chọn duy nhất một tham số là kích thước cửa sổ k để trượt trên tập dữ liệu gốc và thuật toán này không có giải hệ phương trình để lưu hệ số hay phải chọn vị trí các nút để có thể nội suy. Thông qua sử dụng nguyên tắc Sliding Window, thuật toán Moving Average chỉ cần cộng thêm giá trị của dữ liệu mới và loại bỏ đi giá trị cũ nhất nên độ phức tạp sẽ không quá lớn so với Spline vì Spline phải giải hệ số cho từng đa thức thông qua việc giải hệ phương trình và lưu các hệ số đó ở trên ma trận hoặc danh sách. Thêm nữa, giải thuật Moving Average vì theo nguyên tắc “Cửa sổ trượt” cho nên sẽ chỉ thực hiện tính toán ở cục bộ, không bị ảnh hưởng bởi các yếu tố ngoại lai hay những điểm dữ liệu ở xa trong quá khứ.

Nhược điểm của phương pháp Moving Average so với Spline:

- Thuật toán của Moving Average đơn giản và dễ triển khai nhưng nếu xây dựng biểu đồ trên tập dữ liệu gốc thì thuật toán trả về biểu đồ sẽ khó có thể đánh giá chi tiết được dữ liệu và có thể sẽ bỏ qua những chi tiết của dữ liệu gốc đã che giấu. Mặc dù thuật toán cũng xử lý dữ liệu là cục bộ nhưng vì chỉ làm trơn mượt ở mức trung bình, đơn giản theo công thức, chỉ nhìn theo k mẫu dữ liệu hiện tại nên những điểm dữ liệu khác thì Moving Average có thể làm thay đổi, bỏ sót hoặc làm biến dạng xu hướng của dữ liệu hiện tại dẫn đến nếu dùng để đánh giá tổng quan chi tiết hoặc dùng để đánh giá được xu hướng cong của biểu đồ thì phương pháp Moving Average sẽ không đánh giá được kết quả chính xác so với Spline đã và đang thực hiện. Vì thế, phương pháp này ta có thể dùng để đánh giá tổng quan hoặc dữ liệu dài hạn

nhưng chi tiết để nắm bắt rõ xu hướng các giai đoạn thì dùng những phương pháp khác như Spline sẽ cho ra kết quả tốt hơn.



Hình 16: Phương pháp Moving Average khi tăng kích thước cửa sổ

- Khác với Moving Average, cho dù Spline tính toán phức tạp hơn rất nhiều so với Moving Average nhưng điểm đặc biệt là Spline xử lý toàn bộ các điểm dữ liệu gốc thông qua việc chia thành những đa thức nhỏ hơn để xử lý riêng từng điểm dữ liệu. Việc xử lý của Spline được làm thông qua việc chọn các nút nội suy ở trên toàn bộ tập dữ liệu nhằm có thể đánh giá được kết quả tổng thể tốt hơn. Ngoài ra Spline chỉ cần có tập dữ liệu gốc làm tham số để biểu diễn và không có sự tác động của tham số thứ ba cho nên thuật toán sẽ chạy và vẽ đường cong từ đầu đến cuối mà không cần phải thay đổi thuật toán quá nhiều.

Dưới đây là sai số của SMA trên cửa sổ bằng 5 với Spline bậc ba tự nhiên:

- Spline:
 - Sai số tuyệt đối trung bình: $5.783408143923366e - 13$.
 - Sai số bình phương tối thiểu: $2.8461390666329305e - 12$.
- SMA
 - Sai số tuyệt đối trung bình: 1679.832380952381.
 - Sai số bình phương tối thiểu: 4249.78221837532

5 Tài liệu tham khảo

- [1] L. T. Thanh, in *Giáo trình Phương Pháp Tính*, Thành phố Hồ Chí Minh, Đại học Quốc Gia TP.Hồ Chí Minh - Trường Đại Học Bách Khoa, 2023, p. 87.
- [2] T. H. Son, "Nội suy và xấp xỉ hàm," Trường đại học Khoa Học Tự Nhiên - Đại Học Quốc Gia TP.Hồ Chí Minh, Thành phố Hồ Chí Minh, 2025.
- [3] N. Đ. Thúc, "Phương pháp tính," Trường đại học Khoa Học Tự Nhiên - Đại học quốc gia TP.HCM, Thành phố Hồ Chí Minh, 2025.
- [4] Nguyễn Thanh Bình (chủ biên), Đinh Ngọc Thanh, Nguyễn Đình Thúc, Nguyễn Đăng Minh, "Khớp dữ liệu (Data Fitting)," in *Cơ sở toán cho khoa học dữ liệu*, Thành phố Hồ Chí Minh, Đại học Quốc Gia TP.Hồ Chí Minh - Trường đại học Khoa Học Tự Nhiên, 2025, pp. 75-90.
- [5] Cafedev, "Cafedev," 18 12 2020. [Online]. Available: <https://cafedev.vn/tu-hoc-ml-thuc-hien-hoi-quy-da-thuc/>. [Accessed 20 07 2025].
- [6] Wikipedia, "Wikipedia," 31 5 2025. [Online]. Available: https://en.wikipedia.org/wiki/Polynomial_regression. [Accessed 20 07 2025].
- [7] O. J. Msigwa, "MQL5," 06 10 2022. [Online]. Available: <https://www.mql5.com/en/articles/11477>. [Accessed 20 07 2025].
- [8] Wikipedia, "Wikipedia," 05 06 2025. [Online]. Available: https://en.wikipedia.org/wiki/Moving_average. [Accessed 23 07 2025].
- [9] CFI Team, "CFI," [Online]. Available: <https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/simple-moving-average-sma/>. [Accessed 23 07 2025].

- Khánh Linh - Anh Tú, "Lao Động Media," Lao Động, 31 12 2021. [Online].
- [10] Available: <https://laodong.vn/photo/chum-anh-tphcm-2021-nhung-ngay-thang-khong-the-quen-990417.ldo>. [Accessed 22 7 2025].
- Sở Y Tế TP.Hồ Chí Minh trung tâm kiểm soát bệnh tật, "Sở Y Tế TP.Hồ Chí Minh trung tâm kiểm soát bệnh tật," 30 5 2021. [Online]. Available:
- [11] <https://www.hcdc.vn/0-gio-ngay-3152021-thuc-hien-gian-cach-xa-hoi-toan-thanh-pho-ho-chi-minh-5aabdad953634b57ce4a85dda8096cad.html>. [Accessed 22 7 2025].
- Bộ Y Tế, "Bộ Y Tế," 21 12 2021. [Online]. Available: https://moh.gov.vn/tin-tong-hop/-/asset_publisher/k206Q9qkZOqn/content/ngay-21-12-co-16-325-ca-mac-covid-19-ha-noi-lai-nhieu-nhat-ca-nuoc-voi-1-704-ca. [Accessed 22 7 2025].