**TRƯỜNG ĐẠI HỌC THỦ DẦU MỘT**

**KHOA KỸ THUẬT – CÔNG NGHỆ**

**\*\*\*\*\*\*\*\*\*\*\***



**BÁO CÁO TỐT NGHIỆP**

**ĐỀ TÀI:**

**ỨNG DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH**

**ĐỂ MÔ HÌNH HOÁ QUÁ TRÌNH PHÁT TRIỂN**

**CỦA LÁ LAN HỒ ĐIỆP**

**Sinh viên thực hiện : LÊ TRỌNG HẬU**

**Lớp : D16PM01**

**Khoá : 2016-2020**

**Ngành : KỸ THUẬT PHẦN MỀM**

**Giảng viên hướng dẫn : TS. GV. HOÀNG MẠNH HÀ**

**Bình Dương, tháng 07/2020**

**TRƯỜNG ĐẠI HỌC THỦ DẦU MỘT**

**KHOA KỸ THUẬT – CÔNG NGHỆ**

**\*\*\*\*\*\*\*\*\*\*\***



**BÁO CÁO TỐT NGHIỆP**

**ĐỀ TÀI:**

**ỨNG DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH**

**ĐỂ MÔ HÌNH HOÁ QUÁ TRÌNH PHÁT TRIỂN**

**CỦA LÁ LAN HỒ ĐIỆP**

**Sinh viên thực hiện : LÊ TRỌNG HẬU**

**Lớp : D16PM01**

**Khoá : 2016-2020**

**Ngành : KỸ THUẬT PHẦN MỀM**

**Giảng viên hướng dẫn : TS. GV. HOÀNG MẠNH HÀ**

**Bình Dương, tháng 07/2020**

# LỜI CẢM ƠN

Báo cáo tốt nghiệp chuyên ngành Kỹ thuật phần mềm với đề tài “**ỨNG DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH ĐỂ MÔ HÌNH HOÁ SỰ PHÁT TRIỂN CỦA LÁ LAN HỒ ĐIỆP**” là kết quả của quá trình cố gắng không ngừng của bản thân và được sự giúp đỡ, động viên khích lệ của các thầy, bạn bè và người thân. Qua trang viết này tôi xin gửi lời cảm ơn tới những người đã giúp đỡ tôi trong thời gian học tập - nghiên cứu khoa học vừa qua.

Tôi xin tỏ lòng kính trọng và biết ơn sâu sắc đối với thầy TS. Hoàng Mạnh Hà đã trực tiếp tận tình hướng dẫn cũng như cung cấp tài liệu thông tin khoa học cần thiết cho luận văn này.

Xin chân thành cảm ơn Lãnh đạo trường Đại học Thủ Dầu Một, khoa Kỹ thuật – Công nghệ và ngành Kỹ thuật phần mềm đã tạo điều kiện cho tôi hoàn thành tốt học phần môn Báo cáo tốt nghiệp của mình.

Cuối cùng tôi xin chân thành cảm ơn bạn bè, người thân đã giúp đỡ tôi trong quá trình học tập và thực hiện Báo cáo tốt nghiệp.

       Sinh viên thực hiện

Lê Trọng Hậu

# LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng  
dẫn khoa học của TS. Hoàng Mạnh Hà. Các nội dung nghiên cứu, kết quả trong  
đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây.  
Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong luận văn còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm  
về nội dung luận văn của mình.** Trường Đại học Thủ Dầu Một không liên quan  
đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện  
(nếu có).

*Bình Dương, ngày 10 tháng 07 năm 2020*

*Người thực hiện  
 (ký tên và ghi rõ họ tên)*

# NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

Bình Dương, … ngày … tháng… năm 2020

Giáo viên hướng dẫn

(Ký và ghi rõ họ tên)

TS. Hoàng Mạnh Hà

# NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

Bình Dương, … ngày … tháng… năm 2020

Giáo viên hướng dẫn

(Ký và ghi rõ họ tên)

Gv. Hồ Đắc Hưng

# LỜI NÓI ĐẦU

Trong thời đại số ngày nay những gì liên quan đến công nghệ dường như đều thu hút sự chú ý của mọi người. Trong số đó **Ứng dụng Machine Learning** cũng là một trong những nền tảng công nghệ không thể không nhắc đến. Ứng dụng Machine Learning  không ngừng phát triển và đóng vai trò chủ chốt trong nhiều ứng dụng quan trọng. Ứng dụng Machine Learning cung cấp nhiều giải pháp tiềm năng cho nhiều lĩnh vực, đồng thời cũng là một trong những phương tiện đưa chúng ta tiến xa hơn trong thời gian tới như Data Mining (Đào dữ liệu), Natural Language Processing (Tiếp thu ngôn ngữ tự nhiên), và kể cả nhận diện hình ảnh (pattern recognition) và thậm chí expert systems (hệ chuyên gia),…

Trong thực tế ngoài những thứ được liệt kê ở trên Ứng dụng Machine Learning (ML) còn được ứng dụng nhiều hơn như thế. Những ứng dụng của Machine Learning không ngừng phát triển mạnh trên các lĩnh vực liên quan tới công nghiệp, nghiên cứu, dịch vụ,… mà còn phát huy ở các mảng về sản xuất, nông nghiệp. Trên thực tế, nền nông nghiệp ở Việt Nam vẫn đóng vai trò như một trụ đỡ của nền kinh tế, nên việc xây dựng nên một nền nông nghiệp hiện đại, hội nhập quốc tế là một vấn đề cần thực hiện tiên quyết.

Với suy nghĩ đó, và mong được đóng góp một phần nhỏ khả năng của bản thân,

# TÓM TẮT

Mô hình hóa là một hướng nghiên cứu thu hút sự quan tâm của các nhà khoa học vì tính ứng dụng hiệu quả trong thực tế và là lĩnh vực hứa hẹn nhiều thách thức cả trong nghiên cứu cơ bản.

David Mumford, huy chương Field 1966, một nhà toán học chuyên về mô hình hóa tại Đại học Havard, đã chỉ ra rằng tuy Phương trình đạo hàm riêng (PDE) là công cụ tuyệt vời để mô hình hóa một hệ tiến hóa nhưng sử dụng nó lại không đơn giản. Tuy nhiên, có một tiếp cận đơn giản hơn mà vẫn hiệu quả, đó là dùng dữ liệu để xây dựng mô hình thông kê để mô tả quá trình tiến hóa, phát triển của một hệ thống.

Lan Hồ Điệp là loại cây cảnh đem lại lợi ích kinh tế [1]. Để đảm bảo cho Lan phát triển, cần phải theo dõi, kiểm soát, đánh giá sự phát triển của lá [1]. Do vậy, đã có rất nhiều nghiên cứu theo hướng áp dụng các mô hình thống kê để mô hình hóa sự tăng trưởng của lá Lan Hồ Điệp [1,2,3]. Tuy vậy, các công cụ được sử dụng đề là các phần mềm chuyên dụng cho thống kê như SPSS, Stata. Trong đó, quá trình tính toán ra kết quả là các tham số thống kê là bị che dấu như hộp đen đối với người vận hành. Do vậy ta không thể kiểm soát được sai số.

Sau khi tôi nắm được các ý tưởng của Chen trong [1] nhằm mô hình hóa sự phát triển của lá Lan Hồ Điệp bằng mô hình Hồi quy tuyến tính và mô hình hồi quy Logitsic. Tôi đã nhận ra có thể giảm nhỏ hơn sai số tạo ra do cách áp dụng mô hình hồi quy tuyến tính thông qua các phần mềm thống kê.

Việc dùng mô hình tuyến tính để xấp xỉ một đường cong bất kỳ là luôn tạo ra nhiều sai số. Do vậy tôi đã tìm hiểu và đề xuất áp dụng thuật toán LMS để tính ra các hệ số của một mô hình biến đổi tuyến tính bám theo sự thay đổi của dữ liệu thực tế.

Ban đầu mục đích chính của đề tài tốt nghiệp này là giải quyết vấn đề cải thiện độ chính xác cho bài toán của Chen nêu ra trong [1]. Nhưng sau đó kết quả này sẽ được áp dụng tại địa phương, do phải tính lại các tham số trong điều kiện tự nhiên khác so với điều kiện tại các nghiên cứu trong [1].

Để so sánh và đánh giá, tôi đã dùng độ đo RMSE. So sánh trực tiếp với các kết quả trong [1] và được trình bày trong quyển khóa luận này. Kết quả này minh họa tính ưu việt của giải pháp của tôi.

# SUMMARY

Modeling is a research direction that attracts the attention of scientists for its practical applicability and is a field that promises many challenges in basic research.

David Mumford, Field Medal 1966, a mathematician specializing in modeling at Harvard University, points out that while Individual Equation Equation (PDE) is a great tool for modeling an evolutionary system, It's not that simple to use. However, there is a simpler but still effective approach, which is to use data to build a statistical model to describe the evolution and development of a system.

Orchid is an ornamental tree with economic benefits [1]. To ensure Lan's development, it is necessary to monitor, control and evaluate the development of leaves [1]. Therefore, there have been many studies towards applying statistical models to model the growth of Lan Ho Diep leaves [1,2,3]. However, the tools used are specialized software for statistics such as SPSS, Stata. In particular, the calculation process results in statistical parameters that are hidden like a black box to the operator. So we can't control the error.

After I grasped Chen's ideas in [1] in order to model the growth of Lan Ho Diep leaves using the linear regression model and the logitsic regression model. I realized that it is possible to reduce the errors smaller than those created by applying linear regression models through statistical software.

Using a linear model to approximate an arbitrary curve always produces a lot of errors. Therefore, I have studied and proposed the application of the LMS algorithm to calculate the coefficients of a linear transformation model that follow the change of actual data.

At begin, the main purpose of this graduation thesis was to solve the problem of accuracy improvement for Chen's problem mentioned in [1]. But then this result will be applied locally, due to the need to recalculate the parameters in natural conditions different from the conditions in the studies in [1].

For comparison and evaluation, I used the RMSE measure. Direct comparison with the results in [1] and presented in this thesis. This result illustrates the superiority of my solution.

# MỤC LỤC

[LỜI CẢM ƠN ii](#_Toc45698722)

[LỜI CAM ĐOAN iii](#_Toc45698723)

[NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN iv](#_Toc45698724)

[NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN v](#_Toc45698725)

[LỜI NÓI ĐẦU vi](#_Toc45698726)

[TÓM TẮT vii](#_Toc45698727)

[SUMMARY viii](#_Toc45698728)

[MỤC LỤC ix](#_Toc45698729)

[DANH MỤC BẢNG xi](#_Toc45698730)

[DANH MỤC HÌNH xii](#_Toc45698731)

[DANH MỤC THUẬT NGỮ VIẾT TẮT xiii](#_Toc45698732)

[CHƯƠNG 1: BÀI TOÁN MÔ HÌNH HOÁ SỰ TĂNG TRƯỞNG KÍCH THƯỚC LÁ LAN HỒ ĐIỆP 1](#_Toc45698733)

[1.1. LÝ DO THỰC HIỆN ĐỀ TÀI 1](#_Toc45698734)

[1.2. MỤC ĐÍCH NGHIÊN CỨU 2](#_Toc45698735)

[1.4. YÊU CẦU CỦA ĐỀ TÀI 2](#_Toc45698736)

[1.5. Ý NGHĨA KHOA HỌC VÀ Ý NGHĨA THỰC TIỄN CỦA ĐỀ TÀI 3](#_Toc45698737)

[CHƯƠNG 2. ĐỀ XUẤT ÁP DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH CHO BÀI TOÁN MÔ HÌNH HOÁ SỰ TĂNG TRƯỞNG KÍCH THƯỚC LÁ LAN HỒ ĐIỆP 4](#_Toc45698738)

[2.1. MÔ HÌNH HỒI QUY TUYẾN TÍNH. 4](#_Toc45698739)

[**2.1.1.** **Mô hình hồi quy tuyến tính đơn** 4](#_Toc45698740)

[*2.1.1.1.* *Vấn đề mô hình hồi quy* 4](#_Toc45698741)

[*2.1.1.2.* *Ước lượng hệ số hồi quy* 7](#_Toc45698742)

[*2.1.1.3.* *Tính chất ước lượng của các hệ số hồi quy* 9](#_Toc45698743)

[*2.1.1.4.* *Kiểm định giả thuyết* 10](#_Toc45698744)

[**2.1.2.** **Mô hình hồi quy tuyến tính bội** 12](#_Toc45698745)

[*2.1.2.1.* *Phương trình hồi quy* 12](#_Toc45698746)

[*2.1.2.2.* *Ước lượng hệ số hồi quy và tính chất của ƯL* 13](#_Toc45698747)

[*2.1.2.3.* *Kiểm định giả thuyết* 14](#_Toc45698748)

[*2.1.2.4.* *Ước lượng và dự đoán* 16](#_Toc45698749)

[2.2. ÁP DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH CHO BÀI TOÁN MÔ HÌNH HOÁ SỰ PHÁT TRIỂN CỦA LÁ LAN HỒ ĐIỆP. 18](#_Toc45698750)

[**2.2.1.** **Dữ liệu cho mô hình.** 18](#_Toc45698751)

[**2.2.2.** **Thực hiện Hồi quy tuyến tính cho bài toán bằng phần mềm SPSS** 21](#_Toc45698752)

[2.3. MỘT SỐ HẠN CHẾ. 27](#_Toc45698753)

[2.4. NHẬN XÉT 27](#_Toc45698754)

[CHƯƠNG 3. ĐỀ XUẤT ÁP DỤNG THUẬT TOÁN LMS CHO BÀI TOÁN XÁC ĐỊNH CÁC HỆ SỐ CỦA MÔ HÌNH HỒI QUY TUYẾN TÍNH 28](#_Toc45698755)

[3.1. THUẬT TOÁN LMS (LEAST MEAN SQUARE ERROR) 28](#_Toc45698756)

[3.1.1. Giới thiệu chung 28](#_Toc45698757)

[**3.1.2.** **Tổ hợp thích nghi tuyến tính** 28](#_Toc45698758)

[**3.1.3.** **Bài toán xác định ma trận trọng số tối ưu cho tổ hợp thích nghi tuyến tính.** 30](#_Toc45698759)

[3.2. ÁP DỤNG THUẬT TOÁN LMS CHO BÀI TOÁN XÁC ĐỊNH CÁC HỆ SỐ CỦA MÔ HÌNH HỒI QUY TUYẾN TÍNH 35](#_Toc45698760)

[CHƯƠNG 4. KẾT LUẬN 43](#_Toc45698761)

[CHƯƠNG 5: TÀI LIỆU THAM KHẢO 44](#_Toc45698762)

[5.1. TÀI LIỆU SÁCH 44](#_Toc45698763)

[5.2. TÀI LIỆU ONLINE 45](#_Toc45698764)

[PHỤ LỤC 47](#_Toc45698765)

# DANH MỤC BẢNG

[Bảng 2. 1:Độ sạch của oxy ứng với tỷ lệ phần trăm hydrocarbon 4](#_Toc45694915)

[Bảng 2. 2:Phân tích phương sai để kiểm định tính hiệu quả của hồi quy 12](#_Toc45694916)

[Bảng 2. 3: Chỉ số sinh trưởng của lá cây giống HF-1 20](#_Toc45694917)

[Bảng 2. 4: Chỉ số sinh trưởng của lá cây giống HF-3 20](#_Toc45694918)

[Bảng 2. 5: Tóm tắt, phân tích phương sai và phân tích hệ số cho dữ liệu HF-1 23](#_Toc45694919)

[Bảng 2. 6: Các giá trị biến phụ thuộc y và giá trị dự báo y ̂ của dữ liệu HF-1 và HF-3 25](#_Toc45694920)

[Bảng 2. 7: Các giá trị RMSE\_HQTT của mẫu cây HF-1 26](#_Toc45694921)

[Bảng 2. 8: Các giá trị RMSE\_HQTT của mẫu cây HF-3 26](#_Toc45694922)

[Bảng 3. 1: Các giá trị nội suy độ dài lá cây của mẫu cây HF-1 37](#_Toc45694957)

[Bảng 3. 2: Các giá trị nội suy độ dài lá cây của mẫu cây HF-3 38](#_Toc45694958)

# DANH MỤC HÌNH

[Hình 2. 1: Đồ thị rải điểm, đường hồi quy cho số liệu sạch của oxy 6](#_Toc45695307)

[Hình 2. 2: Độ lệch và các đường hồi quy lý thuyết, thực nghiệm. 8](#_Toc45695308)

[Hình 2. 3: Miền biến thiên của các biến hồi quy 17](#_Toc45695309)

[Hình 2. 4:Biểu đồ biểu thị độ dài lá (cm) theo thời gian (day) của các mẫu cây Lan Hồ Điệp dưới các điều kiện chăm sóc riêng 20](#_Toc45695310)

[Hình 2. 5: Cài đặt chạy thuật toán HQTT trên SPSS 22](#_Toc45695311)

[Hình 2. 6: Cài đặt chạy thuật toán HQTT trên SPSS 22](#_Toc45695312)

[Hình 2. 7:Cài đặt chạy thuật toán HQTT trên SPSS 23](#_Toc45695313)

[Hình 3. 1: Tổ hợp thích nghi tuyến tính 29](#_Toc45695317)

[Hình 3. 2: Biểu đồ nội suy các giá trị mẫu cây HF-1 36](#_Toc45695318)

[Hình 3. 3:Biểu đồ nội suy các giá trị mẫu cây HF-3 37](#_Toc45695319)

[Hình 3. 4: Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-1 40](#_Toc45695320)

[Hình 3. 5: Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-1 sau khi đã rút gọn. 40](#_Toc45695321)

[Hình 3. 6: Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-3 41](#_Toc45695322)

[Hình 3. 7: Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-3 sau khi đã rút gọn. 41](#_Toc45695323)

# DANH MỤC THUẬT NGỮ VIẾT TẮT

|  |  |
| --- | --- |
| BNN | Biến ngẫu nhiên |
| LMS | Least Mean Square (Tối thiểu hoá trung bình của bình phương) |
| HQTT | Hồi quy tính tuyến (Linear Regression) |
| MHHQ | Mô hình hồi quy |
| ƯL | Ước lượng |
|  |  |
|  |  |

# CHƯƠNG 1: GIỚI THIỆU CHUNG

## LÝ DO THỰC HIỆN ĐỀ TÀI

Lan Hồ Điệp (*Phalaenopsis*) là một trong những loài lan có hoa lớn, đẹp, bền, có hoa mọc quanh năm, hoa lâu tàn và có nhiều màu sắc. Hiện nay Lan Hồ Điệp nói riêng và các loài hoa lan khác nói chung được xem là cây trồng dem lại hiệu quả kinh tế cao. Do đó cũng đã có nhiều nhà vườn mạnh dạn đầu tư chuyển đổi cơ cấu cây trồng từ lúa, hoa màu sang trồng lan. Hiệu quả kinh tế cao gấp 2-3 lần so với các cây trồng khác. Lan Hồ Điệp hiện nay rất được yêu thích không chỉ về màu sắc kiểu dáng mà còn mang một vẻ đẹp sang trọng và trang nhã. Cùng với sự phát triển nhanh chóng của kinh tế, xã hội, nhu cầu sử dụng hoa nói chung và hoa lan nói riêng ở Việt Nam cũng tăng nhanh. Bên cạnh yêu cầu về số lượng thì chất lượng hoa cũng đòi hỏi ngày càng cao. Hoa lan trên thị trường Việt Nam hiện nay chủ yếu là được nhập về từ Đài Loan, Thái Lan, Trung Quốc… Điều đó cho thấy, quy mô và phương pháp sản xuất hoa Lan Hồ Điệp nói riêng và các loài hoa Lan nói chung ở Việt Nam chưa đáp ứng được nhu cầu của thị trường. Cụ thể là quá trình theo dõi, kỹ thuật chăm sóc còn yếu nên năng suất và chất lượng hoa chưa được đảm bảo.

Các nghiên cứu trong [1] đã chỉ ra những thông số phản ánh sự phát triển của Lan Hồ Điệp mà ta cần theo dõi chặt chẽ. Một trong những thông số đó là độ dài lá [1]. Do vậy, để kịp thời can thiệp khi Lan Hồ Điệp có biểu hiện bất thường ta có mô hình hóa sự tăng trưởng độ dài của lá với đủ tin cậy.

Xuất phát từ nhu cầu thực tế trên, cũng như góp phần cho ngành nuôi trồng lan, tôi áp dụng kỹ thuật, công nghệ thông tin vào nhằm tiến hành nghiên cứu đề tài: “**ỨNG DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH ĐỂ MÔ HÌNH HOÁ SỰ PHÁT TRIỂN CỦA LÁ LAN HỒ ĐIỆP”.**

## MỤC ĐÍCH NGHIÊN CỨU

Mục đích nghiên cứu của đề tài nhằm:

* Nghiên cứu và mối tương quan giữa thời gian sinh trưởng (ngày) và chiều dài của của lá Lan Hồ Điệp.
* Tiến hành mô hình hoá thông qua mô hình hồi quy tuyến tính.
  1. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU
     + Đối tượng nghiên cứu: Tập chung vào nghiên cứu sự phát triển của lá Lan Hồ Điệp, từ đó tìm ra mô hình phát triển tổng quát.
     + Phạm vi nghiên cứu: Do hạn chế về thời gian và tài liệu nghiên cứu, đề tài sẽ tập chung vào sự ảnh hưởng của các yếu tố với cây Lan theo tài liệu NCKH của [Chiachung Chen](https://www.researchgate.net/profile/Chiachung_Chen) và [Meiyu Chien](https://www.researchgate.net/scientific-contributions/2056639897_Meiyu_Chien). [18]
  2. PHƯƠNG PHÁP NGHIÊN CỨU

Để thực hiện đề tài này, người thực hiện sử dụng hai phương pháp nghiên cứu sau:

Phương pháp nghiên cứu lý thuyết:

* Các phương pháp machine learning liên quan tới Linear Regression.
* Các phương pháp nghiên cứu xử lý thuật toán LMS.

Phương pháp nghiên cứu thực nghiệm:

* Thực nghiệm dữ liệu trên SPSS.
* Cài đặt thuật toán trên Matlap làm thực nghiệm.

## Ý NGHĨA KHOA HỌC VÀ Ý NGHĨA THỰC TIỄN CỦA ĐỀ TÀI

* Kết quả của đề tài góp phần hoàn thiện quá trình chăm sóc, nuôi trồng Lan Hồ Điệp làm cơ sở cho sự phát triển trên diện rộng, để có thể ứng dụng vào thực tiễn, sản xuất ở nước ta.
* Phát triển ứng dụng đo kích thước lá Lan Hồ Điệp, từ đó đưa ra các gợi ý, giải pháp cho việc chăm sóc một cách thích hợp.
* Áp dụng thuật toán LMS cho bài toán xác định các tham số thống kê của mô hình hồi quy tuyến tính.
  1. BỐ CỤC CỦA LUẬN VĂN

Bố cục của luận văn bao gồm 4 chương cùng với phần mở đầu, phần mục lục, phần tài liệu tham khảo và phụ lục:

CHƯƠNG 1: GIỚI THIỆU CHUNG

Giới thiệu chung về bố cục của báo cáo bao gồm các phương pháp nghiên cứu và ý nghĩa khoa học thực tiễn và mục đích nghiên cứu đề tài.

CHƯƠNG 2: ĐỀ XUẤT ÁP DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH CHO BÀI TOÁN MÔ HÌNH HOÁ SỰ TĂNG TRƯỞNG KÍCH THƯỚC LÁ LAN HỒ ĐIỆP

Trình bày một số khái niệm, mô hình công thức của mô hình Hồi quy tuyến tính, áp dụng phần mềm SPSS vào việc tính ra mô hình hồi quy.

CHƯƠNG 3: ĐỀ XUẤT ÁP DỤNG THUẬT TOÁN LMS CHO BÀI TOÁN XÁC ĐỊNH CÁC HỆ SỐ CỦA MÔ HÌNH HỒI QUY TUYẾN TÍNH

Trình bày một số khái niệm , mô hình, công thức của thuật toán LMS, viết mã lệnh Matlap để xác định chuỗi các hệ số cần điều chỉnh của mô hình Hồi quy tuyến tính.

CHƯƠNG 4: KẾT LUẬN

Kết luận những mặt tích cực của báo cáo, những hạn chế mà báo cáo chưa đưa ra được, và hướng phát triển.

# CHƯƠNG 2. ĐỀ XUẤT ÁP DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH CHO BÀI TOÁN MÔ HÌNH HOÁ SỰ TĂNG TRƯỞNG KÍCH THƯỚC LÁ LAN HỒ ĐIỆP

## MÔ HÌNH HỒI QUY TUYẾN TÍNH.

* + 1. **Mô hình hồi quy tuyến tính đơn**

Thuật toán Linear Regression (Hồi Quy Tuyến Tính) thuộc nhóm Supervised learning (Học có giám sát) một trong những thuật toán cơ bản nhất của Machine Learning [2]. Hồi quy tuyến tính là một phương pháp rất đơn giản nhưng đã được chứng minh được tính hữu ích cho một số lượng lớn các tình huống.

* + - 1. *Vấn đề mô hình hồi quy*

Nhiều bài toán trong khoa học, kỹ thuật đòi hỏi khảo sát quan hệ giữa hai hoặc nhiều biến. Lấy làm ví dụ, chúng ta xét số liệu ở bảng 2.1, ở đó **y** chỉ thị độ sạch của oxy sinh ra trong quá trình chưng cất hoá học, còn **x** là nồng độ phần trăm của hydrocarbon có mặt bình ngưng bộ phận chưng cất.

Bảng 2. :Độ sạch của oxy ứng với tỷ lệ phần trăm hydrocarbon

|  |  |  |
| --- | --- | --- |
| TT x(%) y(%)  1 0.99 90.01  2 1.02 89.05  3 1.15 91.43  4 1.29 93.74    5 1.46 96.73  6 1.36 94.45  7 0.87 87.59 | TT x(%) y(%)  8 1.23 91.77  9 1.55 99.42  10 1.4 93.65  11 1.19 93.54  12 1.15 92.52  13 0.98 90.56  14 1.01 89.54 | TT x(%) y(%)  15 1.11 89.85  16 1.2 90.39  17 1.26 93.25  18 1.32 93.41  19 1.43 94.98  20 0.95 87.33  21 1.32 94.01 |

Khi thể hiện các điểm (xi, yi) lên đồ thị, ta nhận được đồ thị rải điểm như hình 2.1. Ta nhận thấy, mặc dù không có đường cong đơn giản nào đi qua các điểm này, song có thể khẳng định rằng, các điểm ấy dường như nằm phân tán quanh một đường thẳng với phương trình y = f(x) nào đó. Vậy có thể giả thiết rằng giá trị trung bình của Y – biến chỉ độ sạch khi nồng độ phần tram X của hydrocarbon tại mức x thoả mãn quan hệ

**E (Y | x) = f(x)** (2.1)

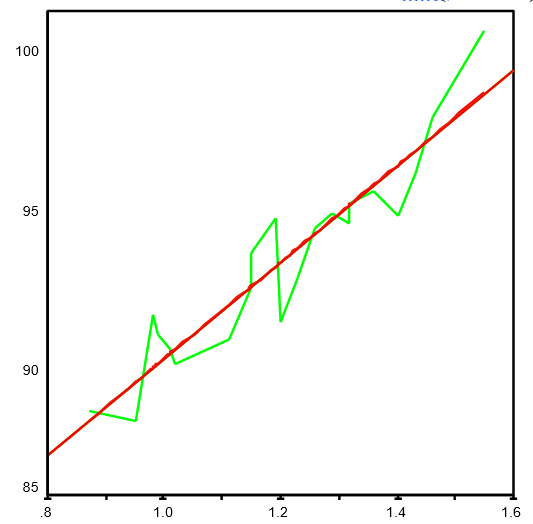
Để tổng quát hoá, chúng ta nên dùng mô hình xác suất bằng cách coi Y là BNN (biến ngẫu nhiên) mà ứng với giá trị x của biến X thì

**Y = f(x) + ε**  (2.2)

với ε là sai lầm ngẫu nhiên.

Trước hết chúng ta xét trường hợp đơn giản nhất, cũng rất hay xảy ra trong thực tế, khi f(x) = ax + b, khi đó (2.2) trở thành

**Y = ax + b + ε**  (2.3)



Hình 2. : Đồ thị rải điểm, đường hồi quy cho số liệu sạch của oxy

Mô hình (2.3) được gọi là mô hình hồi quy (MHHQ) tuyến tính đơn, x được gọi là biến hồi quy ( hay biến độc lập, biến giải thích ), Y được gọi là biến phản hồi (hay biến phụ thuộc, biến được giải thích), a, b được gọi là các tham số hồi quy, a: hệ số chặn, b: hệ số góc: đường thẳng y= ax+b được gọi là đường **hồi quy**.

Mô hình được gọi là **tuyến tính** là vì nó tuyến tính với các tham số a, b (a, b có luỹ thừa 1); được gọi là **đơn** vì có một biến hồi quy.

Giả sử ở quan sát i biến X nhận giá trị xi , biến Y nhận giá trị yi và sai lầm ngẫu nhiên là εi. Như vậy, dưới dạng quan sát, mô hình (2.3) trở thành

|  |  |
| --- | --- |
| **{** | **y1=ax1+b+ε1** |
| **. . . . . . . . .** | (2.4) |
| **yn=axn+b+εn** |

với yi là các BNN.

Để khảo sát mô hình, chúng ta phải tiến hành các thí nghiệm, các phép đo đạc hay các phép quan sát, gọi chung là quan sát, để có bộ số liệu {(xi, yi)}. Thông qua bộ số liệu này, người ta đưa ra các xấp xỉ (ước lượng) tốt cho các tham số. Mô hình với các hệ số đã ước lượng được gọi là mô hình thực nghiệm (empirical model) hay mô hình lọc (filted model). Dùng mô hình thực nghiệm chúng ta có thể tiến hành một số dự đoán, tính các giá trị cực trị cũng như các khía cạnh của vấn đề điều khiển.

* + - 1. *Ước lượng hệ số hồi quy*

Bây giờ, giả sử các BNN y1,…,yn nhận các giá trị cụ thể nào đó, vẫn ký hiệu là y1,…,yn . Khi đó

**εi = yi – (axi+b)** (2.5)

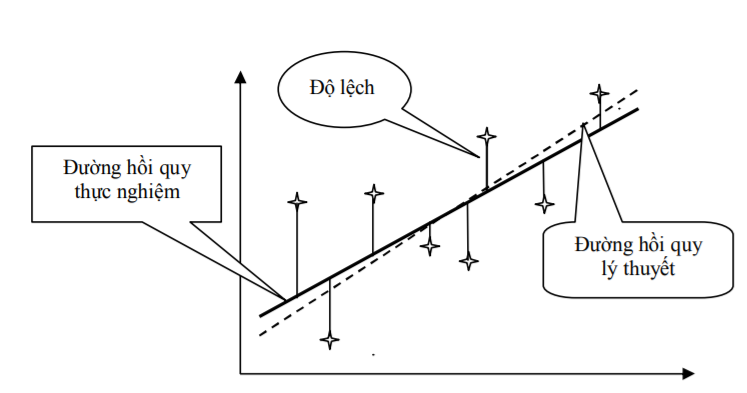
thể hiện độ lệch của quan sát thứ i so với đường hồi quy lý thuyết (hình 2.2). Tổng bình phương các độ lệch

thể hiện “chất lượng” của việc xấp xỉ số liệu bởi đường hồi quy lý thuyết. Ta không thể biết đường hồi quy lý thuyết, việc ta có thể làm là tìm các hệ số a, b để

(2.6)

Vì λ(a,b) là đa thức bậc 2 của 2 ẩn a, b; điều kiện cần để nó đạt cực tiểu là

(2.7)



Hình 2. : Độ lệch và các đường hồi quy lý thuyết, thực nghiệm.

Đây là hệ 2 phương trình tuyến tính bậc nhất của a, b. Ta có thể tính được nghiệm của hệ này là:

(2.8)

Trong đó:

**; ; ;** . (2.9)

Với các ƯL này ta được phương trình hồi quy thực nghiệm

**.** (2.10)

Phương pháp tìm các ước lượng xấp xỉ của hệ số như trên gọi là phương pháp bình phương cực tiểu.

Các phương trình (2.5) - (2.10) áp dụng với mọi giá trị cụ thể của các BNN y1 ,..., yn nên chúng cũng đúng cho các BNN này. Dưới đây, khi áp dụng các phương trình này và khi không sợ lầm lẫn, ta không phân biệt các BNN y1 ,..., yn với các giá trị cụ thể của chúng.

* + - 1. *Tính chất ước lượng của các hệ số hồi quy*

Từ (2.8) ta có được . Như vậy, đường hồi quy đi qua điểm “trung tâm” () của số liệu. ƯL hệ số (2.8) hoàn toàn không cần các giả thiết về các thành phần ngẫu nhiên εi . Để có các tính chất tốt của ƯL, cần có những giả thiết đặt lên các thành phần ngẫu nhiên này. Giả thiết dễ chấp nhận là chúng có kỳ vọng không, cùng phương sai σ2, độc lập; giả thiết tiếp sau là chúng có phân bố chuẩn:

ε1,…,εn độc lập, cùng phân bố chuẩn N(0; σ2). (2.11)

Khi đó ƯL hệ số có những tính chất thống kê tốt thể hiện ở định lý sau.

**Định lý 1**. Khi điều kiện (2.11) thỏa mãn thì:

* lần lượt là ƯL không chêch của tham số a và b:

**; E[**  (2.12)

* Phương sai của các ƯL và được tính như sau

**, ,** (2.13)

* ƯL không chệch của phương sai chung của mô hình cho bở

(2.14)

Với

: dự báo quan sát thứ i.

: phần dư thứ i.

Định nghĩa. Đối với mô hình HQTT đơn, sai số chuẩn hoá ( thực nghiệm) của hệ số góc và hệ số chặn lần lượt được xác định bởi

**;** (2.15)

Trong đó, được tính theo (2.14)

* + - 1. *Kiểm định giả thuyết*

Một khâu quan trọng để kiểm tra tính phù hợp của MHHQ là kiểm định giả thuyết. Các hệ số và những BNN nên có thể làm một số kiểm định về chúng. Ta luôn nhớ rằng điều kiện (2.11) phải được thỏa mãn. Các đối thuyết đưa ra dưới đây đều là 2 phía. Độc giả có thể đưa ra đối thuyết 1 phía với điều chỉnh thích hợp các ngưỡng phê phán [3].

1. Sử dụng kiểm định *T*

Hệ số góc là tham số quan trọng nhất của MHHQ tuyến tính đơn. Xét bài toán kiểm định giả thuyết hai phía:

**H0: b =b0/H1: b b0.**  (2.16)

Ở đây, b0 là giá trị cho trước. Từ giả thiết (2.11), yi là các BNN độc lập và yi ~ N(a + bxi; σ2 ). là tổ hợp tuyến tính của các BNN yi  nên nó cũng có phân bố chuẩn. Theo Định lý 1, có phân bố chuẩn N(b; / Sxx). Ngoài ra, như trong chứng minh của Định lý trên, (n - 2) / có phân bố khi bình phương với n - 2 bậc tự do và độc lập với .

**.**  (2.17)

Như vậy, chúng ta sẽ báo bỏ H0 ( ở mức ý nghĩa ) nếu

**.**  (2.18)

Trường hợp đặc biệt quan trọng là khi b0=0;

**H0: b = 0/H1: b 0.**  (2.19)

Điều này liên quan đến ý nghĩa (hay tác dụng) của hồi quy (significance of regression): Nếu không bác bỏ H0 (coi b=0) thì có nghĩa rằng không có một quan hệ tuyến tính nào giữa X và Y (có thể là quan hệ thực sự của X và Y là quan hệ phi tuyến), sự thay đổi của biến X không kéo theo sự thay đổi dự đoán biến Y, X không có (hoặc rất ít) tác dụng để dự đoán Y; dự đoán cho Y tốt nhất nên dùng.

Tương tự, giả thuyết liên quan đến hệ số chặn là

**H0: a = a0/H1: a 0.**  (2.20)

Bởi vì

**.** (2.21)

Nên giả thuyết bị bác bỏ ở mức nếu

**.** (2.22)

1. Phân tích phương sai

Phương pháp phân tích phương sai được dùng để kiểm định tính hiệu quả của việc lập mô hình. Trước hết, từ chỗ, bình phương hai vế rồi lấy tổng ta được:

**= +** (2.23)

Chúng ta xác định các đại lượng sau đây:

Tổng bình phương đầy đủ**:**

**SST = SYY =** ,

Tổng bình phương hồi quy:

**SSR =** ,

Tổng bình phương các phần dư ( các sai số):

**SSE =** (2.24)

Biểu thức (2.23) được viết lại:

**🡺 SST=SSR+SSE** (2.23’)

Có thể chứng minh rằng, **SSR/** và **SSE/** là những BNN độc lập, có phân bố khi bình phương với 1 và n – 2 bậc tự do tương ứng. Như vậy, nếu giả thuyết H0 : b = 0 là đúng thì

**F0 =**  (2.25)

Có phân bố **F(1;n-2).**

Các đại lượng MSR, MSE gọi chung là bình phương trung bình. Nói chung, bình phương trung bình được tính bằng cách lấy tổng bình phương chia cho bậc tự do của nó.

Chúng ta sẽ bác bỏ H0 nếu **F0 > (1;n-2).**

Trong các phần mềm thống kê, thủ tục kiểm định được trình bày ở bảng phân tích phương sai giống như Bảng 2.2.

Bảng 2. :Phân tích phương sai để kiểm định tính hiệu quả của hồi quy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nguồn | Tổng các bình phương | Bậc tự do | Bình phương trung bình | F0 | P – giá trị |
| Hồi quy  Sai số  Đầy đủ | SSR  SSE  SST | 1  n-2  n-1 | MSR  MSE |  | P |

Nếu P - giá trị lớn hơn mức ý nghĩa chọn trước, chúng ta phải chấp nhận giả thuyết b = 0 , tức là việc xây dựng mô hình không có tác dụng. Cần phải tìm mô hình khác, lấy thêm số liệu…

***Lưu ý*.** Chứng minh được, thủ tục phân tích phương sai và thủ tục kiểm định T cho bài toán kiểm định giả thuyết 2 phía (2.16) là tương đương theo nghĩa chấp nhận giả thuyết hay bác bỏ giả thuyết là đồng thời với 2 thủ tục này. Tuy nhiên, kiểm định T linh động hơn, có thể xét kiểm định 1 phía, trong khi phân tích phương sai chỉ có thể xét 1 phía.

* + 1. **Mô hình hồi quy tuyến tính bội**

MHHQ tuyến tính bội là sự mở rộng tự nhiên của MHHQ tuyến tính đơn.

* + - 1. *Phương trình hồi quy*

1. Dạng quan sát và dạng ma trận

Giả sử mối quan hệ giữa biến phụ thuộc (biến phản hồi) Y và k biến độc lập (biến hồi quy) cho bởi mô hình

**.** (2.26)

Trong đó là các tham số chưa biết, gọi là các hệ số hồi quy, gọi là hệ số chặn, là các hệ số góc, là sai số ngẫu nhiên có kỳ vọng 0 và phương sai .

Ta có thể viết ngắn gọn (2.26) thành

(2.27)

Hoặc

(2.28)

1. Tuyến tính hoá một số mô hình

Mô hình (2.28) vì nó tuyến tính với các tham số .

Trong ứng dụng chúng ta thường gặp mô hình dạng

(2.29)

Trong đó , là các hàm của biến hồi quy .

Đây là mô hình tuyến tính với các tham số , phi tuyến với các biến Xét phép biến đổi

**.**

Ta có thể đưa (2.29) về dạng

(2.30)

Là mô hình tuyến tính với cả tham số lẫn các biến hồi quy.

* + - 1. *Ước lượng hệ số hồi quy và tính chất của ƯL*

Giả thiết đầu tiên cần có là ma trận X có số hang ít nhất bằng số cột, p = k+1 n, và hạng của nó bằng số cột:

**Rank(X) = p** (2.31)

Khi đó, ƯL làm cực tiểu tổng bình phương các sai số

Gọi là ƯL bình phương cực tiểu, ký hiệu là , cho bởi:

**=**  (2.32)

ƯL cho sai số chung của mô hình là

(2.33)

Với

: dự báo quan sát thứ i.

: phần dư thứ i.

Nhận thấy vế phải của (2.33) có chứa mẫu số n - p . Vậy, khi số biến hồi quy p tăng lên, (chẳng hạn với hồi quy đa thức, khi số bậc của đa thức tăng) có thể sai số mô hình tăng lên. Ta sẽ có mô hình cực tồi nếu p ≅ n.

Để nghiên cứu các tính chất của ƯL tham số, giống với trường hợp có 1 biến hồi quy, cần có giả thiết:

**ε1 ,…, εn độc lập, cùng phân bố chuẩn N(0; σ2 ) .** (2.34)

**Định lý 2:** Với giả thuyết (2.31), (2.33) thì:

1. là ƯL không chệch của véc tơ tham số : .
2. Ma trận covarian của cho bởi:

Cov( =

1. theo (2.2.8) là ƯL không chệch của :

.

* + - 1. *Kiểm định giả thuyết*

1. Kiểm định ý nghĩa của hồi quy.

Đó là kiểm tra xem có một quan hệ tuyến tính nào đó giữa biến phản hồi Y với một tập con nào đó của các biến hồi quy x1 ,..., xk hay không. Cụ thể là xét bài toán kiểm định:

**: = = … = / :**

với ít nhất một j.

Nếu bị bác bỏ thì có nghĩa là ít ra một trong các biến hồi quy x1,…,xk có ý nghĩa đối với mô hình.

Dưới giả thuyết H0 có thể chứng minh tổng bình phương hồi quy và tổng bình phương các sai số theo (2.24) là những BNN độc lập và có bậc tự do tương ứng là k và n - p . Thế thì

**.** (2.35)

Từ đó giả thuyết bị bác bỏ ở mức α nếu F0 ≥ fα (k; n − p).

Các phần mềm thường dùng P – giá trị và đưa ra bảng phân tích phương sai cho thủ tục vừa nêu. Người ta cũng xét kiểm định cho một tập con của các hệ số β0, β1,..., βk bằng 0.

1. Hệ số xác định bội R2 và hệ số xác định hiệu chỉnh

Với mô hình hồi quy nhiều biến định nghĩa hệ số xác định bội R2 và các tính chất của nó như với trường hợp hồi quy đơn:

**.**

Tính chất đặc biệt của hệ số xác định là nó không giảm khi tăng số biến hồi quy. Từ đó, hệ số xác định khó nói cho ta biết việc tăng biến có lợi gì hay không, nhất là khi sự gia tăng hệ số xác định là nhỏ. Vì thế nhiều nhà phân tích lại thích dùng hệ số xác định hiệu chỉnh (adjusted R2 ):

**.**  (2.36)

Mẫu ở vế phải là hằng số, còn tử là ước lượng của sai số; nó bé nhất khi và chỉ khi hệ số xác định hiệu chỉnh lớn nhất. Từ đó, một quy tắc lựa chọn biến hồi quy là:

Chọn một số trong các biến hồi quy x1 ,..,xk để lớn nhất.

1. Kiểm định một tham số triệt tiêu (kiểm định T)

Xét bài toán kiểm định một tham số đơn lẻ nào đó triệt tiêu:

**: = / : ( j.**

Nếu giả thuyết không bị bác bỏ thì có nghĩa rằng biến hồi quy tương ứng không bị loại khỏi mô hình. Thống kê kiểm định là:

**Tj= = .**  (2.37)

Trong đó Cjj là phần tử thứ j của đường chéo chính của ma trận C=(X’X)-1 tương ứng với .

Vì **Tj ~ T(n-p)** nên giả thuyết bị bác bỏ nếu **.**

* + - 1. *Ước lượng và dự đoán*

1. Khoảng tin cậy cho tham số đơn lẻ

Khoảng tin cậy 100(1 )% − α cho tham số j β cho bởi

**, =**. (2.38)

1. Khoảng tin cậy cho đáp ứng trung bình

Giả sử quan sát tương lai thực hiện tại mức x01,...,x0k của các biến hồi quy x1 ,..., xk . Đặt x0 = (1, x01 , ...,x0k )T . Đáp ứng trung bình tại điểm này là , ƯL điểm của nó là:

**.**

Đối với MHHQ tuyến tính bội. khoảng tin cậy 100(1-)% cho đáp ứng trung bình tại điểm x01,…,x0k là

(2.39)

1. Dự đoán cho quan sát mới.

ƯL điểm của dự đoán cho quan sát tương lai tại mức x01,…,x0k của các biến độc lập là

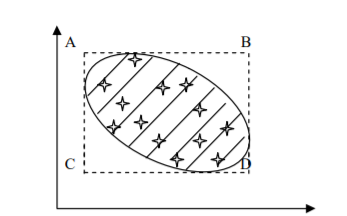
**.**

Khoảng dự đoán 100(1-)% cho quan sát tương lai này là

. (2.40)

1. Vấn đề ngoại suy với mô hình hồi quy bội

Vẫn có những chú ý tương tự như với hồi quy đơn, song vấn đề cần thận trọng hơn. Chẳng hạn, với mô hình có hai biến hồi quy x, y miền biến thiên của các biến hồi quy ở Hình 2.3 phải hiểu là elip chứ không phải hình chữ nhật ABCD. Tình hình sẽ khó khăn hơn khi số biến hồi quy tăng lên.



Hình 2. : Miền biến thiên của các biến hồi quy

* + - 1. **Phân tích phần dư**

Với mô hình bội, người ta cũng tiến hành lập đồ thị phần dư chuẩn hóa di = ei / như với mô hình đơn. Nếu có không quá 95% các giá trị i d nằm trong dải (-2; 2) và phần dư có dáng điệu tương đối đều đặn quanh trục hoành thì chấp nhận mô hình. Trái lại, phải tiến hành phân tích lại.

Tuy nhiên, có hai điểm khác biệt.

*Thứ nhất*, ngoài lập đồ thị phần dư chuẩn hóa theo thời gian (theo chỉ số i), theo chiều tăng của một vài biến hồi quy xi nào đó, theo chiều tăng của dự báo , khi xét mô hình với một nhóm con các biến hồi quy, người ta còn lập đồ thị phần dư theo biến hồi quy chưa tham gia vào mô hình. Nếu phát hiện ra đồ thị phần dư chuẩn hóa theo biến này không đạt yêu cầu thì có nhiều khả năng biến hồi quy đó cần phải tham gia vào mô hình.

*Thứ hai*, thay cho đồ thị phần dư chuẩn hóa di, người ta thấy rằng đồ thị phần dư điều chỉnh ri (còn gọi là phần dư student hóa – (studentized residual)) ưu việt hơn, trong đó

(2.41)

Với là phần tử chéo thứ i của ma trận

**H = X(XTX)-1XT**.

*Lưu ý:* **0 < hii 1 => di < ri**

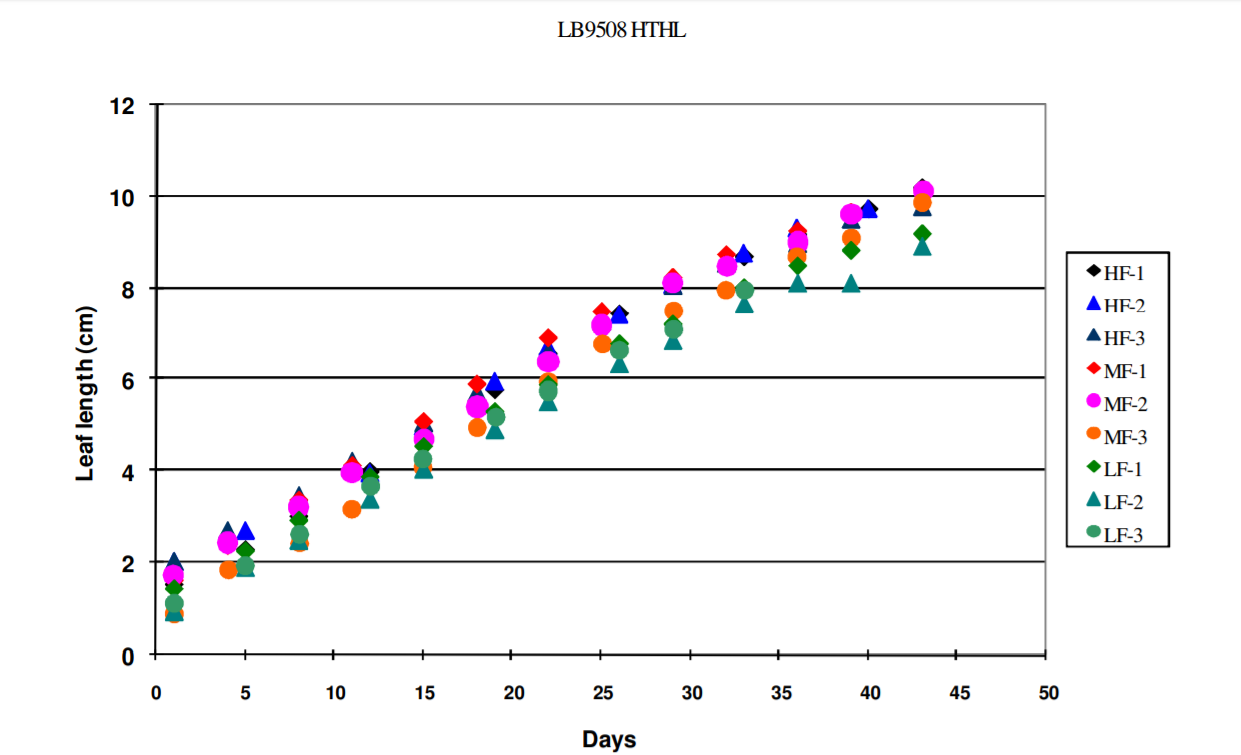
## ÁP DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH CHO BÀI TOÁN MÔ HÌNH HOÁ SỰ PHÁT TRIỂN CỦA LÁ LAN HỒ ĐIỆP.

* + 1. **Dữ liệu cho mô hình.**

Lan Hồ Điệp đã trở thành một trong những loại cây trồng trong chậu có giá trị cao nhất trên thế giới. Để đảm bảo số lượng và chất lượng sản xuất hoa lan, thì các vấn đề ảnh hưởng đến các yếu tố tăng trưởng thực vật, như nhiệt độ, cường độ ánh sáng và nồng độ phân bón cần được hiểu rõ [1].

Lá cây khoẻ mạnh là một trong những yêu cầu cơ bản cho việc canh tác Lan Hồ Điệp. Quá trình quang hợp và thoát hơi nước có liên quan đến sự phát triển của lá. Sự phát triển của lá liên quan đến truyền năng lượng, tích luỹ khối lượng và vận chuyển chất. Sự phát triển của chiều dài lá bị ảnh hưởng bởi nhiệt độ, cường độ ánh sáng, quản lý thụ phấn, …và các yếu tố khác. Các điều kiện tối ưu của các yếu tố này rất quan trọng đối với người trồng lan, nhằm tạo ra một biểu đồ phát triển tốt nhất, giúp cải thiện và điều chỉnh tỉ lệ, phương pháp chăm sóc[1].

Mẫu nghiên cứu sự phát triển lá của các giống mẫu hoa Lan Hồ Điệp ( theo số liệu nghiên cứu từ thạc sĩ Chiachung Chen và Meiyu Chien, Đại học Quốc gia ChungHsing, Đài Loan) [18] trên các cây khác nhau dưới các điều kiện chăm sóc khác nhau về được quan sát và ghi nhận lại khoảng thời gian 45 ngày.



Hình 2. :Biểu đồ biểu thị độ dài lá (cm) theo thời gian (day) của các mẫu cây Lan Hồ Điệp dưới các điều kiện chăm sóc riêng

Ta phân tách dữ liệu và chọn chỉ số sinh trưởng của lá cây có mức tỉ lệ phân bón cao, nhiệt độ ngày cao và cường độ ánh sáng cao: 333 mgL-1, 26-30, 200-280 mol m-2 s-1 (HF-1 và HF-3)

Bảng 2. : Chỉ số sinh trưởng của lá cây giống HF-1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x (ngày) | 1 | 5 | 8 | 12 | 19 | 22 | 26 | 33 | 40 | 43 |
| y (độ dài) | 1.5 | 2.3 | 3 | 4 | 5.75 | 6.75 | 7.5 | 8.75 | 9.75 | 10.2 |

Bảng 2. : Chỉ số sinh trưởng của lá cây giống HF-3

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x (ngày) | 1 | 4 | 8 | 11 | 15 | 18 | 22 | 28 | 36 | 38 | 43 |
| y (độ dài) | 2 | 2.75 | 3.5 | 4.2 | 5 | 5.7 | 6.75 | 8 | 9.25 | 9.5 | 9.7 |

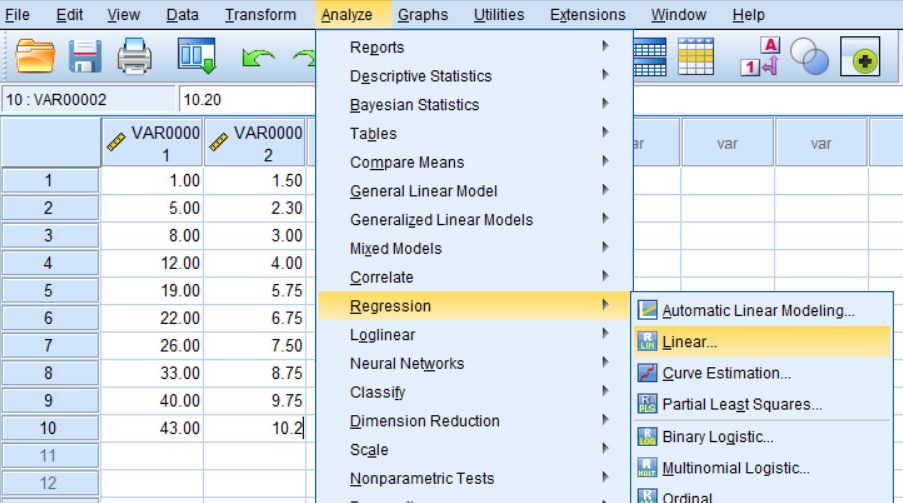
* + 1. **Thực hiện Hồi quy tuyến tính cho bài toán bằng phần mềm SPSS**

Các phần mềm thống kê ngày nay cho phép phân tích mô hình với số biến hồi quy lên đến hàng ngàn và số quan sát lên đến hàng chục vạn. Chúng ta cần có những kiến thức cơ bản để tận dụng những lợi thế của các phần mềm này. Mỗi phần mềm có những thế mạnh của nó, song chúng đều có phần phân tích hệ số và phân tích phương sai.

Ở bài toán mô hình hoá này, ta dùng phần mềm “*IBM* ***SPSS*** *Statistics Data Editor (v.25)”* để xử lí dữ liệu.

SPSS (viết tắt của Statistical Package for the Social Sciences) là một chương trình máy tính phục vụ công tác thống kê. Phần mềm SPSS hỗ trợ xử lý và phân tích dữ liệu sơ cấp - là các thông tin được thu thập trực tiếp từ đối tượng nghiên cứu, thường được sử dụng rộng rãi trong các các nghiên cứu điều tra, phân tích xã hội học và kinh tế lượng.

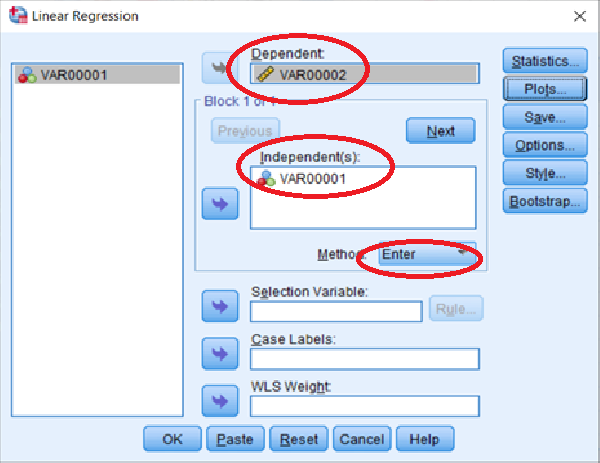
Chọn chương trình phân tích: **Analyze / Regression / Linear…**



Hình 2. : Cài đặt chạy thuật toán HQTT trên SPSS

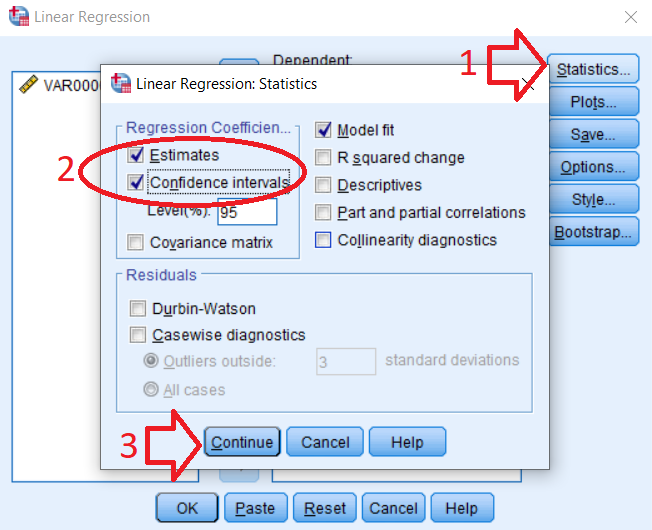
Chọn biến: Từ danh sách các biến, đẩy biến y sang ô biến phụ thuộc (**Dependent**), đẩy biến x sang ô biến độc lập ( **Independent**).

Phương pháp lọc mô hình: Trong ô phương pháp (**Method**) ta chọn enter. Phần mềm sẽ lọc mô hình có tất cả các biến.



Hình 2. : Cài đặt chạy thuật toán HQTT trên SPSS

Tìm ƯL cho tham số và khoảng tin cậy của chúng: **Statistics** / chọn **Estimates** và **Confidence** **intervals** / **Continue**.



Hình 2. :Cài đặt chạy thuật toán HQTT trên SPSS

Chạy chương trình: **OK**.

Kết quả sau khi chạy chương trình:

Bảng 2. : Tóm tắt, phân tích phương sai và phân tích hệ số cho dữ liệu HF-1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
| Variables Entered | | | Variables Removed | | | Method | | | |
| VAR00001b | | | . | | | Enter | | | |
| . | | | | | | | | |
| R | R Square | | | Adjusted R Square | | | Std. Error of the Estimate | | | |
| .994a | .988 | | | .987 | | | .36473 | | | |
|  |  | | |  | | |  | | | |
|  | | Sum of Squares | | | df | | | Mean Square | | | F | Sig. |
| Regression | | 87.991 | | | 1 | | | 87.991 | | | 661.432 | .000b |
| Residual | | 1.064 | | | 8 | | | .133 | | |  |  |
| Total | | 89.055 | | | 9 | | |  | | |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | |
| B | Std. Error | Beta | Lower Bound | Upper Bound |
| (Constant) | 1.482 | .209 |  | 7.104 | .000 | 1.001 | 1.962 |
| VAR00001 | .214 | .008 | .994 | 25.718 | .000 | .195 | .233 |

Ta thấy hệ số xác định R2 = 0.988, vậy có 98.8% số liệu được giải thích bởi mô hình; Đây là một tỉ lệ khá lớn. ƯL cho phương sai của mô hình là . Mức ý nghĩa của thống kê F là 0.000, rất nhỏ so với 0.01: Mô hình có tác dụng tốt dể giải thích số liệu. Tất cả các mức ý nghĩa của thống kê T của các tham số đều nhỏ hơn 0.05 (giá trị cực đại ứng với hằng số = 0.000 ). Như vậy, các kiểm định T không bác bỏ mô hình. Mô hình dự tuyển là:

(2.42)

Với mô hình hồi quy 2.42 vừa tìm được, ta tiến hành phân tích phương sai, tìm ra **căn bậc hai trung bình bình phương sai số** (RMSE) để kiểm định tính hiệu quả của mô hình hồi quy.

Bằng cách lần lượt thay các giá trị x ( giá trị ngày trong dữ liệu mẫu ) trong hai bộ dữ liệu vào mô hình hồi quy 2.42 ta thu được các giá trị (giá trị dự báo ) tương ứng.

Bảng 2. : Các giá trị biến phụ thuộc y và giá trị dự báo y ̂ của dữ liệu HF-1 và HF-3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **HF-1** | | |  | **HF-3** | | |
| **x** | **y (thực tế)** | **(giá trị dự đoán)** | **x** | **y (thực tế)** | **(giá trị dự đoán)** |
| 1 | 1.5 | 1.696 | 1 | 2 | 1.696 |
| 5 | 2.3 | 2.552 | 4 | 2.75 | 2.338 |
| 8 | 3 | 3.194 | 8 | 3.5 | 3.194 |
| 12 | 4 | 4.05 | 11 | 4.2 | 3.836 |
| 19w | 5.75 | 5.548 | 15 | 5 | 4.692 |
| 22 | 6.75 | 6.19 | 18 | 5.7 | 5.334 |
| 26 | 7.5 | 7.046 | 22 | 6.75 | 6.19 |
| 33 | 8.75 | 8.544 | 28 | 8 | 7.474 |
| 40 | 9.75 | 10.042 | 36 | 9.25 | 9.186 |
| 43 | 10.2 | 10.684 | 38 | 9.5 | 9.614 |
|  | | | 43 | 9.7 | 10.684 |

Ứng với mẫu cây HF-1 , theo 2.23 về tổng bình phương sai số, với L=3,ta được:

RMSE\_HQTT(1) =

= 0.046519

RMSE\_HQTT(2) =

Thực hiện tương tự, ta tìm được chuỗi các giá trị RMSE\_HQTT của mẫu HF-1 và HF-3:

Bảng 2. : Các giá trị RMSE\_HQTT của mẫu cây HF-1

|  |  |
| --- | --- |
| RMSE\_HQTT(1) | 0.046519 |
| RMSE\_HQTT(2) | 0.034547 |
| RMSE\_HQTT(3) | 0.02698 |
| RMSE\_HQTT(4) | 0.118968 |
| RMSE\_HQTT(5) | 0.18684 |
| RMSE\_HQTT(6) | 0.187384 |
| RMSE\_HQTT(7) | 0.111272 |
| RMSE\_HQTT(8) | 0.120652 |

Bảng 2. : Các giá trị RMSE\_HQTT của mẫu cây HF-3

|  |  |
| --- | --- |
| RMSE\_HQTT(1) | 0.118599 |
| RMSE\_HQTT (2) | 0.131959 |
| RMSE\_HQTT (3) | 0.106999 |
| RMSE\_HQTT (4) | 0.120439 |
| RMSE\_HQTT (5) | 0.180807 |
| RMSE\_HQTT (6) | 0.241411 |
| RMSE\_HQTT (7) | 0.198124 |
| RMSE\_HQTT (8) | 0.097923 |
| RMSE\_HQTT (9) | 0.328449 |

Sau khi quan sát chuỗi các giá trị RMSE của mô hình HQTT, ta thấy chúng không có hướng hội tụ (tiến dần) về 0.

Vậy câu hỏi đặt ra, là có phương pháp nào khác, mà khi áp dụng vào có thể cho giá trị sai số nhỏ hơn hay không?

## MỘT SỐ HẠN CHẾ.

Phần mềm SPSS ngoài những ưu điểm như có bộ soạn thảo tương tự excel, cho phép nhập vào các dữ liệu mô tả các thuật toán của chúng, chính vì vậy khá dễ sử dụng [16], giao diện cho phép người và máy tương tác đơn giản để tạo ra các đồ thị,... thì cũng có những nhược điểm có thể thấy khi thực hiện bài toán HQTT như:

* Sự phân bố dữ liệu không đồng nhất.
* Đối với các bài toán có ít dữ liệu, độ chính xác sẽ bị giảm.
* Người dùng không thể can thiệp vào quá trình xử lí của thuật toán.
* Chỉ số chuỗi các sai số (RMSE) vẫn cao, khó để tiến dần về 0.
  1. NHẬN XÉT

Vì về mặt bản chất, mô hình hoá bằng HQTT là dùng phương trình đường thẳng để xấp xỉ hoá tập dữ liệu sự phát triển của lá Lan hồ điệp.

Quan sát trên hình (2.4) ta nhận thấy rằng độ đo RMSE có thể giảm khi ta thay thế việc xấp xỉ đường cong (bám theo phân bố của dữ liệu).

Để tìm cách cải thiện, khắc phục những điều này, ta thử áp dụng phương pháp **áp dụng LMS vào mô hình Hồi quy tuyến tính.**

# CHƯƠNG 3. ĐỀ XUẤT ÁP DỤNG THUẬT TOÁN LMS CHO BÀI TOÁN XÁC ĐỊNH CÁC HỆ SỐ CỦA MÔ HÌNH HỒI QUY TUYẾN TÍNH

Như đã đề cập, ở chương 2, việc xấp xỉ đường cong sẽ giảm sai số RMSE. Tôi đã tham khảo các phương pháp xấp xỉ bằng đường cong như [ ], tuy vậy các phương pháp này có nhược điểm …

* 1. THUẬT TOÁN LMS (LEAST MEAN SQUARE ERROR)

### Giới thiệu chung

Thuật toán bình phương nhỏ nhất (LMS) là một lớp bộ lọc thích ứng được sử dụng để mô phỏng bộ lọc mong muốn bằng cách tìm các hệ số bộ lọc liên quan đến việc tạo ra bình phương trung bình nhỏ nhất của tín hiệu lỗi (chênh lệch giữa tín hiệu mong muốn và tín hiệu thực tế). Đó là một phương pháp giảm độ dốc ngẫu nhiên trong đó bộ lọc chỉ được điều chỉnh dựa trên lỗi tại thời điểm hiện tại.

Nó được phát minh vào năm 1960 bởi giáo sư Đại học Stanford, Bernard Widrow và bằng tiến sĩ đầu tiên của ông. Sinh viên, Ted Hoff. [17]

* + 1. **Tổ hợp thích nghi tuyến tính**

Trong các hệ thống xử lý tín hiệu thích nghi, tổ hợp thich nghi tuyên tính (Hình ) là phần tử cơ bản, đóng vai trò quan trọng.



Hình 3. : Tổ hợp thích nghi tuyến tính

Các mô hình thích nghi được ứng dụng nhiều trên thực tế như: dự đoán, sửa sóng, nhận dạng và lọc nhiễu đều sử dụng các phiên bản của tổ hợp thích nghi tuyến tính. Thuật toán LMS cũng được xây dựng để làm việc trên tổ hợp thích nghi tuyến tính. Cấu tạo chính bao gồm:

* Tín hiệu đầu vào với các phần tử trong đó là biến thời gian rời rạc. thường được mô tả qua hàm xung Dirac như sau:

**,**

trong đó:

**,**

là độ lớn, hoặc biên độ của .

Với , có thể coi là tín hiệu được làm trễ lần của .

* Các trọng số có thể điều chỉnh được .
* Bộ lấy tổng số học.
* Một đầu lấy tín hiệu ra với

***.***  (3.1)

Các trọng số là các số thực và được điều chỉnh sao cho giống nhất với (xem hình 3.1). Việc điều chỉnh các trọng số được coi là thủ tục thích nghi cho tổ hợp. Theo công thức 3.1, đầu ra là kêt quả tổ hợp tuyến tính từ các thành phần tín hiệu đầu vào.

* + 1. **Bài toán xác định ma trận trọng số tối ưu cho tổ hợp thích nghi tuyến tính.**

Để mô hình hoá thủ tục thích nghi, ta sử dụng các ký hiệu sau:

* là vector trọng số của tổ hợp thích nghi tuyến tính, với

**,** (3.2)

là vector có độ dài .

* là tín hiệu huấn luyện mà ta biết trước, trong đó là biến thời gian rời rạc.
* là vector tín hiệu đầu vào tại thời điểm, với

, (3.3)

trong đó:

: Tín hiệu đầu vào tại thời điểm ,

: Tín hiệu được làm trễ 1 lần từ tín hiệu ,

: Tín hiệu được làm trễ lần từ tín hiệu .

là vector có độ dài .

Bài toán đặt ra là ta muốn nhận được với sao cho giống nhất với tín hiệu huấn luyện theo nghĩa đạt được một cực tiểu. Việc điều chỉnh các trọng số được coi như là bài toán xác định các phần tử với , sao cho khi sử dụng công thức (3.1), đầu ra giống nhất với .

Lời giải cho bài toán này dựa trên giả thiết rằng tín hiệu và là tín hiệu dừng theo nghĩa dưới đây.

**Định nghĩa 1.** *Dãy dừng là dãy ngẫu nhiên có phân phối xác suất không thay đổi theo thời gian.*

Nếu dãy là dừng và có biểu diễn hàm phân phối của tại thời điểm , , …, thì ta có :

*,*

với *.*

Trong xử lý tín hiệu, tín hiệu dừng được coi như dạng yếu của dãy dừng (xem [7]). Ta có định nghĩa tín hiệu dừng như sau:

**Định nghĩa 2**. *Tín hiệu dừng là tín hiệu ngẫu nhiên có moment bậc 1 và 2 không thay đổi theo thời gian.*

Nếu tín hiệu là tín hiệu dừng thì

, với .

,

với .

Bài toán xác định ma trận trọng số tối ưu cho tổ hợp thích nghi tuyến tính yêu cầu và là tín hiệu dừng theo định nghĩa 2 và tìm sao cho đạt cực tiểu. Nếu mỗi khi moment bậc 1 và moment bậc 2 của chúng chuyển thành giá trị cố định khác, ta phải tìm lại ma trận trọng số tối ưu khác, sao cho tại đó lại đạt cực tiểu. Sau khi tính được , nếu các thông số của không thay đổi thì cũng không thay đổi.

Các kết quả đạt được được trình bày như sau :

Các kết quả ban đầu được N. Wiener đưa ra trong [9\*]. Tuy chưa giải quyết hoàn toàn được yêu đầu đặt ra nhưng đã là cơ sở quan trọng để những tác giả sau đó phát triển đúng hướng. Lời giải như sau: Nếu hàm để cực tiểu hóa được chọn là , thì

.

Ma trận trọng số được tính như sau:

, (3.4)

trong đó là ma trận trọng số tối ưu, tại đó đạt giá trị cực tiểu,

, (3.5)

và

. (3.6)

Với giả thiết và là tín hiệu dừng như trên, các phần tử của và đều là hằng số.

Tuy nhiên, lời giải trên của N. Wiener chưa giải quyết được trường hợp sau: Mỗi khi các thông số thống kê của của tín hiệu hoặc tín hiệu huấn luyện thay đổi, dẫn đến các phần tử của ma trận và thay đổi, chuyển thành các hằng số khác. Do vậy, một cách tự động phải được tính lại theo các công thức (3.5), (3.6) và (3.4). Nhưng trong thực tế, ta chỉ có được bằng cách đo tín hiệu đầu vào dưới dạng số thực tại những thời điểm chứ không xác định được tín hiệu dưới dạng hàm số nào đó. Do vậy không thể biết được khi nào thay đổi để tính lại . Vậy lời giải Wiener trong công thức (3.4) chưa đủ để thực hiện bộ lọc thích nghi.

* + 1. **Thuật toán thích nghi của B. Widrow**

Để giải quyết vấn đề nêu trên, B. Widrow đã đề xuất thay đổi cách tiếp cận, chuyển từ tính trực tiếp thành tìm kiếm ngẫu nhiên bằng phương pháp lặp [4]. Khi đó bài toán được mô tả tương đương:

Cho công thức lặp tính ma trận trọng số tối ưu như sau :

,

trong đó

: Ma trận trọng số tại thời điểm .

: Ma trận trọng số tại thời điểm *,*

: Tham số hiệu chỉnh tại thời điểm .

Bài toán đặt ra là :

* Tìm Biểu thức tính sao cho mỗi khi đạt đến , thì tham sô hiệu chỉnh và không thay đổi nếu tính chất dừng của và cũng không thay đổi. Nếu tính chất dừng thay đổi, để bắt đầu lại các bước lặp tìm ma trận tối ưu khác.
* Xác định điều kiện để có khả năng hội tụ đến .

Các kết quả đạt được trong [4] được trình bày như sau:

Tham số hiệu chỉnh được B. Widrow và học trò của ông là T. Hoff đề xuất cho thuật toán LMS :

,

trong đó

 : Kích thước bước, quyết định tốc độ hội tụ và độ ổn định của thuật toán. Dải ổn định của thuật toán được xác định trong công thức sau

. (3.7)

Thuật toán LMS có ưu điểm nổi bật về tính đơn giản trong tính toán vì được tính chỉ từ đầu vào và đầu ra của tổ hợp tuyến tính (Hình 3.1).

Từ các kết quả trên, thuật toán LMS được tóm tắt như sau :

* Nạp tín hiệu huấn luyện .
* Gán .
* Gán giá trị khởi tạo cho các phần tử của .
* Chọn để thuật toán ổn định.
* Lặp
  + Gán.
  + Tính .
  + Tính .
  + Cho đến khi hết dãy tín hiệu thì dừng lại.

Với việc chuyển thành bài toán tìm kiếm ngẫu nhiên như trên, thuật toán LMS có khả năng tìm được đối với lớp bài toán không phải là tín hiệu dừng, khi đó không còn dạng bậc 2, là trường hợp không áp dụng được lời giải Wiener. Tuy nhiên điều kiện ổn định trong công thức (3.7) chưa chính xác khi sử dụng thuật toán LMS trong các mô hình được phát triển từ tổ hợp thích nghi tuyến tính như bộ lọc nhiễu Adaptive FIR, đặc biệt là Adaptive IIR. Cần phải xác định thêm điều kiện cần và đủ cho sự hội tụ của thuật toán LMS. Đây cũng là vấn đề có thể được giải quyết trong phần  **‘Áp dụng LMS cho bài toán xác định các hệ số của mô hình hồi quy tuyến tính ‘ .**

* 1. ÁP DỤNG THUẬT TOÁN LMS CHO BÀI TOÁN XÁC ĐỊNH CÁC HỆ SỐ CỦA MÔ HÌNH HỒI QUY TUYẾN TÍNH

Xét thấy dữ liệu mẫu quá ít (10 giá trị) để có thể áp dụng vào mô hình HQTT và cho ra một kể quả chính xác, ta tiến hành phân tích nâng số lượng giá trị dữ liệu bằng phương pháp **nội suy.**

**Nội suy** là phương pháp ước tính giá trị của các điểm dữ liệu chưa biết trong phạm vi của một [tập hợp](https://vi.wikipedia.org/wiki/T%E1%BA%ADp_h%E1%BB%A3p" \o "Tập hợp) rời rạc chứa một số điểm dữ liệu đã biết.

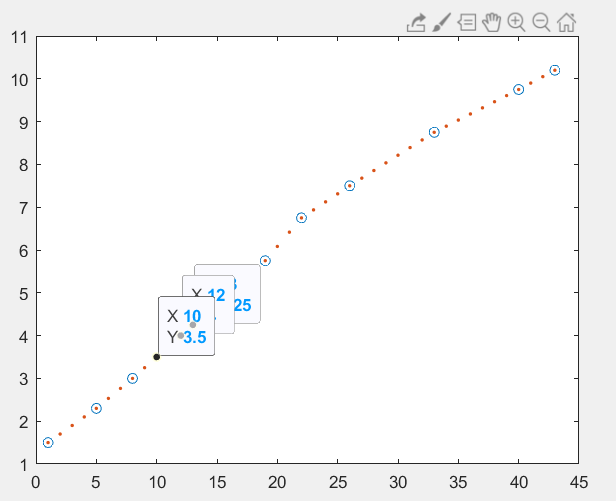
Ví dụ, chúng ta có một số giá trị của hàm số chưa biết *f* như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | 0 | 1 | 2 | 3 | 4 |
| *f(x)* | 0 | 0.5 | 3 | 9 | 17 |

Nội suy là phương thức để ước tính hàm  *f* tại các điểm trung gian, chẳng hạn như điểm x = 2,5.

Có rất nhiều phương pháp nội suy khác nhau, ở bài toán này, ta sử dụng phương pháp **Nội suy tuyến tính (Linear)** bằng phần mềm Matlap

Ở đây, từ phương pháp nội suy ta tìm các giá trị nằm trong khoảng từ ngày 1 tới ngày 43, biểu đồ nội suy nhận được:



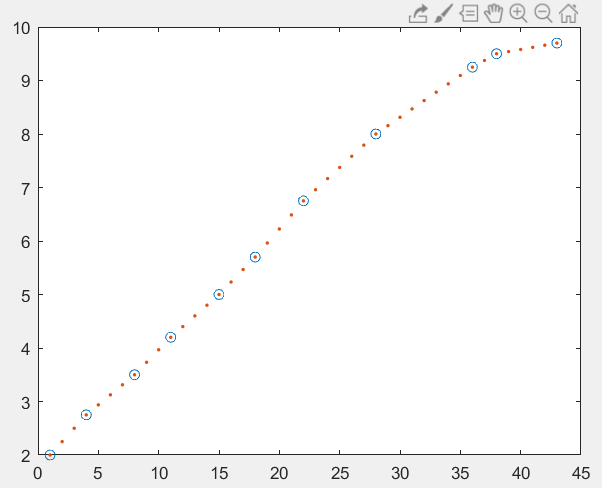
Hình 3. : Biểu đồ nội suy các giá trị mẫu cây HF-1

Ta thu được bảng dữ liệu chi tiết:

Bảng 3. : Các giá trị nội suy độ dài lá cây của mẫu cây HF-1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | y | 8 | 3 | 16 | 5 | 24 | 7.125 | 32 | 8.5741 | 40 | 9.75 |
| 1 | 1.5 | 9 | 3.25 | 17 | 5.25 | 25 | 7.3125 | 33 | 8.75 | 41 | 9.9 |
| 2 | 1.7 | 10 | 3.5 | 18 | 5.5 | 26 | 7.5 | 34 | 8.8929 | 42 | 10.05 |
| 3 | 1.9 | 11 | 3.75 | 19 | 5.75 | 27 | 7.6786 | 35 | 9.0357 | 43 | 10.2 |
| 4 | 2.1 | 12 | 4 | 20 | 6.0833 | 28 | 7.8 | 36 | 9.1786 |  |  |
| 5 | 2.3 | 13 | 4.25 | 21 | 6.4167 | 29 | 8.0357 | 37 | 9.3214 |  |  |
| 6 | 2.533 | 14 | 4.5 | 22 | 6.75 | 30 | 8.2143 | 38 | 9.4643 |  |  |
| 7 | 2.7667 | 15 | 4.75 | 23 | 6.9375 | 31 | 8.3929 | 39 | 9.6071 |  |  |

Tương tự, ta thay dữ liệu giá trị mẫu với mẫu cây HF-3, thu được biểu đồ nội suy và bảng dữ liệu:



Hình 3. :Biểu đồ nội suy các giá trị mẫu cây HF-3

Và bảng dữ liệu chi tiết:

Bảng 3. : Các giá trị nội suy độ dài lá cây của mẫu cây HF-3

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | y | 8 | 3.5 | 16 | 5.2333 | 24 | 7.1667 | 32 | 8.625 | 40 | 9.58 |
| 1 | 1.52 | 9 | 3.7333 | 17 | 5.4667 | 25 | 7.375 | 33 | 8.7813 | 41 | 9.62 |
| 2 | 2.25 | 10 | 3.9667 | 18 | 5.7 | 26 | 7.5833 | 34 | 8.9375 | 42 | 9.66 |
| 3 | 2.5 | 11 | 4.2 | 19 | 5.9625 | 27 | 7.7917 | 35 | 9.0938 | 43 | 9.7 |
| 4 | 2.75 | 12 | 4.4 | 20 | 6.225 | 28 | 8 | 36 | 9.25 |
| 5 | 2.9375 | 13 | 4.6 | 21 | 6.4875 | 29 | 88.1563 | 37 | 9.375 |
| 6 | 3.125 | 14 | 4.8 | 22 | 6.75 | 30 | 8.3125 | 38 | 9.5 |
| 7 | 3.3125 | 15 | 5 | 23 | 6.9583 | 31 | 8.4688 | 39 | 9.54 |

Với phương pháp nội suy như trên, ta có thể nhân số lượng phần tử trong một bộ dữ liệu lên gấp nhiều lần, nhằm tăng độ chính xác khi áp dụng thuật toán.

Thực hiện viết mã lệnh trên phần mềm Matlap để biểu diễn đồ thị các chỉ số của các RMSE dùng trong HQTT và RMSE khi áp dụng LMS.

|  |
| --- |
| function [d,sstbbpkc,W ] = LMS( X,Y,mu,L )  W(1)= 0.5;  for k = 1:length(Y)  tg=X(1,k);  yhat(1,k)=tg\*W(k);  d(1,k)=Y(1,k) - yhat(1,k) ;  W(k+1) = W(k) + 2\*mu\*d(1,k)\*tg;  end;  for k=1:(length(Y)-L)  for m=1:L  y1(1,m)=Y(1,k+m);  ymu(1,m)=yhat(1,k+m);  end;  [sstbbpkc(k)]=rmse( y1,ymu);  end;  end |

Lưu đồ thuật toán:

W(1) = 0.5

Input x(n)

Tính

Tính

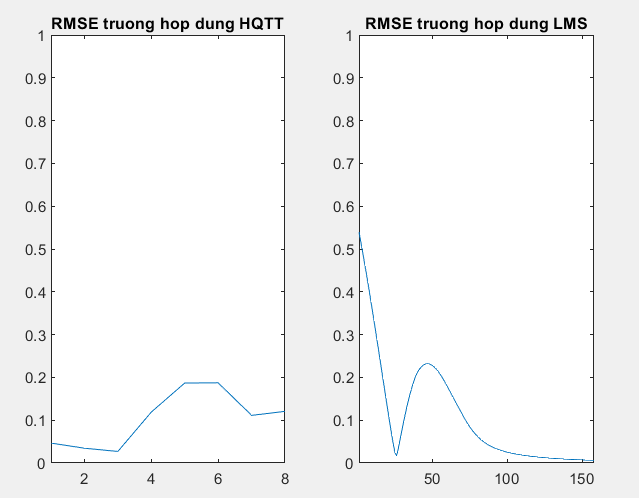
Tính 2

Tính w(n+1) = w(n) +

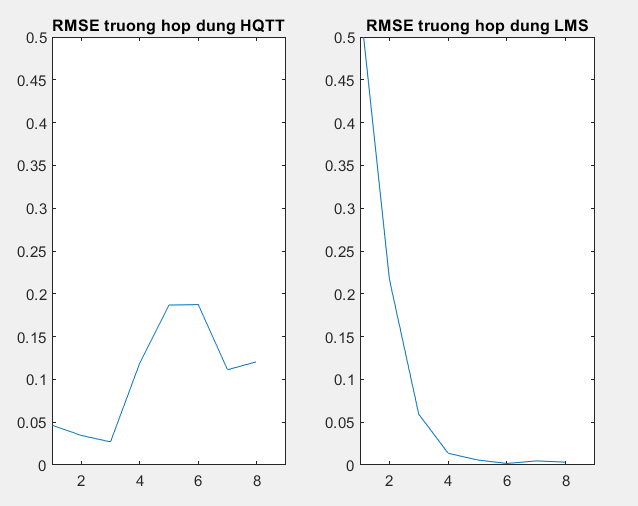
Y

N

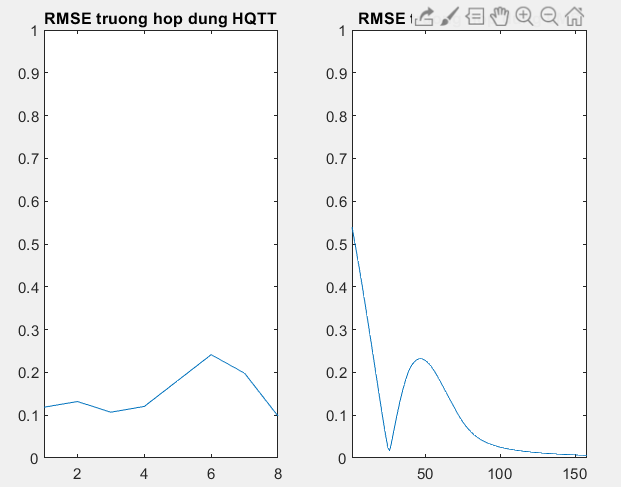
Kết quả biểu đồ của thuật toán chuển qua mã lệnh khi chạy trên Matlap:



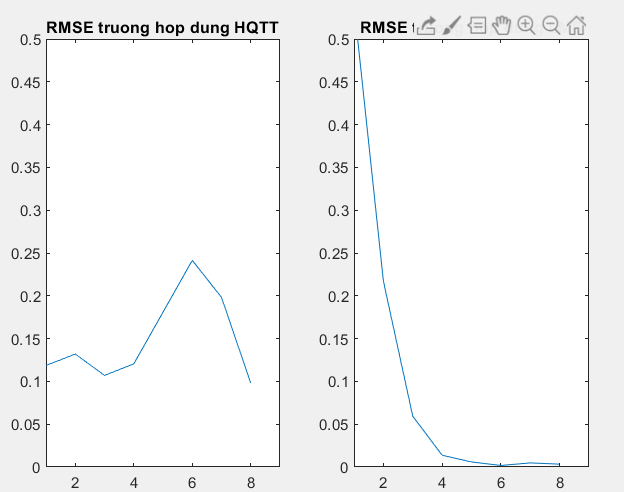
Hình 3. : Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-1



Hình 3. : Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-1 sau khi đã rút gọn.



Hình 3. : Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-3



Hình 3. : Đồ thị biểu diễn RMSE trường hợp HQTT và trường hợp LMS của dữ liệu HF-3 sau khi đã rút gọn.

Quan sát trên hình ta thấy rằng các giá trị RMSE khi áp dụng thuật toán LMS đều nhỏ hơn so với trường hợp áp dụng mô hình Hối quy tuyến tính.

Như vậy, bài toán nêu ra ở **Chương 2.** đã được chứng minh.

# CHƯƠNG 4. KẾT LUẬN

* 1. **ĐÁNH GIÁ KẾT QUẢ**

Bài báo cáo đã tìm ra được mô hình hồi quy đẻ mô hình hoá sự phát triển của lá Lan Hồ Điệp, là tiền đề để phát triển các mô hình, ứng dụng hỗ trợ quá trình theo dõi, chăm sóc khả năng sinh trưởng của cây hoa.

Tìm hiểu về mô hình hồi quy tuyến tính cũng như thuật toán LMS, là những kiến thức không được học trong chương trình bậc đại học của người thực hiện đề tài.

Làm thực nghiệm để so sánh, đánh giá giá trị RMSE khi áp dụng thuật toán LMS so với khi sử dụng mô hình hồi quy, kết quả cho thấy chuỗi giá trị RMSE trên LMS có giảm đi đáng kể, tạo nên độ chính xác cao hơn.

Bên cạnh đó, vẫn còn những điểm cần phải khắc phục. Vì mặt thời gian tìm hiểu, nghiên cứu còn hạn chế, nên người thực hiện đề tài chưa đưa ra được sản phẩm như website hay ứng dụng phục vụ cho việc đo đạc, phân tích. Kiến thức nền tảng về học máy (0machine learning) nói chung và hồi quy tuyến tính ( Linear Regression) nói riêng vẫn còn chưa hoàn thiện.

* 1. **HƯỚNG PHÁT TRIỂN**

Đề xuất hướng phát triển cho đề tài: Phát triển thành sản phẩm, phần mềm như website hoặc ứng dụng di động chuyên dụng, có tích hợp chức năng đo trực tiếp bằng camera giám sát được đặt tại vườn ươm.

# TÀI LIỆU THAM KHẢO

* 1. **TÀI LIỆU SÁCH**

[1] Anthura BY (2005). Cultivation Guide *Phalaenopsis* – Knowledge for Professionals. The Netherlands.

[2] Vũ Hữu Tiệp (2019), Machine Learning cơ bản, NXB Khoa học Kỹ thuật

[3] Trần Mạnh Tuấn. Xác suất & Thống kê - Lý thuyết và thực hành tính toán. Nhà Xuất bản Đại học Quốc gia Hà Nội 2004

[4] B. Widrow, Samuel D. Stearns, (1985), *Adaptive Signal Processing*, Englewood Cliffs, Prentice-Hall.

[5] Daniel Olguín Olguín, Frantz Bouchereau, Sergio Martínez, “*Adaptive Notch Filter for EEG Signals Based on the LMS Algorithm with Variable Step-Size Parameter*” . Proceedings of the Conference on Information Sciences and Systems, The John Hopkins University, March 16-18, 2005

[6] B. Widrow, ‘*Adaptive Signal Processing* ‘ (Englewood Cliffs, NJ 07632: Prentice-Hall, 1985).

[7] Robert M. Gray and Lee D. Davisson, (2009), *An Introduction to Statistical Signal Processing*, Cambridge University Press.

[8] Hoang Manh Ha “***VARIABLE STEP SIZE LMS FILTER FOR ECG SIGNALS* ”** Proceedings of the International Conference on BioMedical Engineering, The Hanoi University of Technology, July 27-29, 2007

[9\*] N. Wiener, (1949), *Extrapolation and Smoothing of Stationary time series, with Engineering Aplications*, Wiley.

* 1. **TÀI LIỆU ONLINE**

[9] <https://caohock24.files.wordpress.com/2012/11/q5-chuong-5-2010-cho-cao-hoc-cntt-doc.pdf>

[10] <https://viblo.asia/p/linear-regression-hoi-quy-tuyen-tinh-trong-machine-learning-4P856akRlY3>

[11] <https://machinelearningmastery.com/simple-linear-regression-tutorial-for-machine-learning/>

[12] <http://www.favri.org.vn/images/Tap_chi/190405-tc12116.pdf>

[13] <https://vjst.vn/vjsta/view-file/1176>

[14] <http://iasvn.org/upload/files/YE1Y7KYGD7KQNC%20-%20Nguyen%20Thi%20Kim%20Ly.pdf>

[15] <https://vi.wikipedia.org/wiki/N%E1%BB%99i_suy#cite_note-2>

[16] <https://ttmhue81.blogspot.com/2019/09/uu-iem-va-nhuoc-iem-cua-phan-mem-spss.html>

[17] <https://en.wikipedia.org/wiki/Least_mean_squares_filter>

[18]<https://www.researchgate.net/publication/267261550_The_leaf_growth_model_and_influencing_factors_in_Phalaenopsis_orchid>

# PHỤ LỤC

**MÃ CHƯƠNG TRÌNH THỰC NGHIỆM**

Ngôn ngữ: Matlap

## **Phụ lục 1: Chương trình chính (noisuy2.m)**

clear all; close all;

x=[1 5 8 12 19 22 26 33];

y=[1.5 2.3 3 4 5.75 6.75 7.5 8.75];

X= 1:0.1:33;

Y = interp1(x,y,X);

mu=0.001;

L=3;

[d,sstbbpkc,W] = LMS( X,Y,mu,L );

rmse\_hqtt1=[0.046518667 0.034546667 0.02698 0.118968 0.18684 0.187384 0.111272 0.120652];

subplot(1,2,1); plot(rmse\_hqtt1); title('RMSE truong hop dung HQTT.'); axis([1 8 0 1]);

subplot(1,2,2); plot(sstbbpkc); title('RMSE truong hop dung LMS'); axis([1 158 0 1]);

for k=1:length(rmse\_hqtt1)

rutgon(k)=sstbbpkc(40\*k - 39)

end;

figure,

subplot(1,2,1); plot(1:8,rmse\_hqtt1); title('RMSE truong hop dung HQTT.'); axis([1 9 0 0.5]);

subplot(1,2,2); plot(1:8,rutgon); title('RMSE truong hop dung LMS'); axis([1 9 0 0.5]);

%đoạn mã trong comment dùng khi muốn hiển thị data HF-3

% rmse\_hqtt2=[0.118598667 0.131958667 0.106998667 0.120438667 0.180806667 0.241410667 0.198124 0.097922667

]

% subplot(1,2,1); plot(rmse\_hqtt2); title('RMSE truong hop dung HQTT..'); axis([1 8 0 1]);

% subplot(1,2,2); plot(sstbbpkc); title('RMSE truong hop dung LMS'); axis([1 158 0 1]);

%

% for k=1:length(rmse\_hqtt2)

% rutgon(k)=sstbbpkc(40\*k - 39)

% end;

% figure,

% subplot(1,2,1); plot(1:8,rmse\_hqtt2); title('RMSE truong hop dung HQTT..'); axis([1 9 0 0.5]);

% subplot(1,2,2); plot(1:8,rutgon); title('RMSE truong hop dung LMS'); axis([1 9 0 0.5]);

## **Phụ lục 2: Hàm tính trọng số bằng thuật toán LMS (LMS.m)**

function [d,sstbbpkc,W ] = LMS( X,Y,mu,L )

W(1)= 0.5;

for k = 1:length(Y)

tg=X(1,k);

yhat(1,k)=tg\*W(k);

d(1,k)=Y(1,k) - yhat(1,k) ;

W(k+1) = W(k) + 2\*mu\*d(1,k)\*tg;

end;

for k=1:(length(Y)-L)

for m=1:L

y1(1,m)=Y(1,k+m);

ymu(1,m)=yhat(1,k+m);

end;

[sstbbpkc(k)]=rmse( y1,ymu);

end;

end

## **Phụ lục 3: Hàm tính sai số trung bình bình phương khai căn (rmse.m)**

function [ kq ] = rmse( y1,yhat )

tong=0;

dodai=length(y1);

for k=1:dodai

tong=tong+(y1(k)-yhat(k))^2;

end;

can=sqrt(tong);

kq=can/dodai;

end