

ณัฐกร นวรัตน : การวิเคราะห์ตัวแบบเชิงบวกนัยทั่วไปคู่ (ANALYSIS OF DOUBLE GENERALIZED ADDITIVE MODELS) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.เจษฎา ตัณฑนุช, 128 หน้า.

คำสำคัญ: ตัวแบบเชิงบวกนัยทั่วไปคู่/การวิเคราะห์เชิงทำนาย/การเรียนรู้ของเครื่อง

วิทยานิพนธ์นี้ต้องการพัฒนาและขยายแนวคิดการสร้างตัวแบบทางคณิตศาสตร์เพื่อใช้ในการทำนายไปสู่ตัวแบบเชิงบวกนัยทั่วไปคู่ เนื้อหาวิทยานิพนธ์ได้แสดงให้เห็นถึงตัวแบบเชิงทำนายต่าง ๆ ได้แก่ ตัวแบบเชิงเส้น ตัวแบบเชิงเส้นนัยทั่วไป ตัวแบบเชิงเส้นนัยทั่วไปสองชั้น รวมไปถึงตัวแบบเชิงบวกนัยทั่วไป และการประมาณค่าพารามิเตอร์ต่าง ๆ ในการสร้างตัวแบบ ผลของการศึกษาแสดงให้เห็นถึงโครงสร้างทางคณิตศาสตร์ของตัวแบบเชิงบวกนัยทั่วไปคู่ ได้แก่ ส่วนประกอบสุม โครงสร้างค่าเฉลี่ย โครงสร้างค่าการกระจาย ฟังก์ชันเชื่อมโยง และการประมาณค่าพารามิเตอร์ของตัวแบบดังกล่าวด้วยวิธี เสมือนพหุนามเชิงแบบฉบับ และการประมาณโดยใช้เทคนิคการเรียนรู้ของเครื่องอีก 4 แบบ ได้แก่ เอ็กซ์เพลนเนเบิลบูสติงแมชชีน เอ็กซ์ตรีมเกรเดียนต์บูสติง ปาสุ่ม และซัพพอร์ตเวกเตอร์รีเกรสชัน พบว่าตัวแบบเชิงบวกนัยทั่วไปคู่แตกต่างจากตัวแบบเชิงบวกนัยทั่วไปโดยไม่เพียงแต่จะเป็นการสร้างตัวแบบของค่าเฉลี่ย แต่มีการสร้างตัวแบบของค่าการกระจายอีกด้วย เพื่อให้เห็นถึงประสิทธิภาพการทำนายของตัวแบบที่พัฒนาขึ้น ได้นำไปทดสอบการสร้างตัวแบบกับชุดข้อมูล 4 ชุด ได้แก่ Fetal Health Monitoring Dataset, Life Style Data, Health Checkup Result Dataset และ Life Expectancy (WHO) Fixed Dataset ในแต่ละข้อมูลทำการแบ่งเป็นชุดเรียนรู้ ชุดปรับพารามิเตอร์ และ ชุดทดสอบ 3 กลุ่มได้แก่ 60:20:20, 70:15:15 และ 80:10:10 และทำการเปรียบเทียบประสิทธิภาพกับตัวแบบเชิงบวกนัยทั่วไปด้วยเครื่องวัดต่าง ๆ ได้แก่ เอไอซี ค่าควรจะเป็นลบ ค่าความคลาดเคลื่อนสมบูรณ์เฉลี่ย รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย สัมประสิทธิ์การกำหนด ซีอาร์พีเอส และพีไอซีพี95 โดยแต่ละชุดข้อมูลจะให้การทำนายตัวแปรลักษณะต่าง ๆ กัน 4 ตัวแปร ทำให้ได้ตัวแบบเชิงบวกนัยทั่วไปทั้งหมด 240 ตัวแบบ และตัวแบบเชิงบวกนัยทั่วไปคู่ทั้งหมด 240 ตัวแบบ ผลการดำเนินการพบว่าเมื่อวัดประสิทธิภาพในเชิงควรจะเป็นและในเชิงความน่าจะเป็น เห็นได้ชัดว่าตัวแบบเชิงบวกนัยทั่วไปคู่มีประสิทธิภาพในการทำนายดีกว่าอย่างชัดเจน

สาขาวิชาคณิตศาสตร์และภูมิสารสนเทศ
ปีการศึกษา 2568

ลายมือชื่อนักศึกษา ณัฐกร
ลายมือชื่ออาจารย์ที่ปรึกษา Dr. Tanthanuch
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม อ.ณัฐกร

NATAKON NAWARATANA : ANALYSIS OF DOUBLE GENERALIZED ADDITIVE MODELS.

THESIS ADVISOR : ASST. PROF. JESSADA TANTHANUCH, Ph.D. 128 PP.

Keyword: DOUBLE GENERALIZED ADDITIVE MODEL, PREDICTIVE ANALYSIS, MACHINE LEARNING

This thesis aims to develop and extend the concept of mathematical modeling for prediction toward the Double Generalized Additive Model (DGAM) framework. The thesis presents a range of predictive models, including linear models, generalized linear models, double generalized linear models, and generalized additive models (GAM), together with methods for parameter estimation in model construction.

The results demonstrate the mathematical structure of the DGAM, which comprises the random component, mean structure, dispersion structure, link functions, and parameter estimation procedures. Parameter estimation is carried out using classical spline-based techniques as well as four machine learning approaches: Explainable Boosting Machine, Extreme Gradient Boosting, Random Forest, and Support Vector Regression. The findings indicate that the DGAM differs fundamentally from the GAM in that it does not solely model the mean structure but additionally incorporates a dispersion model.

To evaluate the predictive performance of the proposed models, experiments are conducted on four datasets: the Fetal Health Monitoring Dataset, Life Style Data, Health Checkup Result Dataset, and the Life Expectancy (WHO) Fixed Dataset. For each dataset, the data are partitioned into training, validation, and testing sets using three split ratios: 60:20:20, 70:15:15, and 80:10:10. The predictive performance of the proposed models is compared with that of the generalized additive model using multiple evaluation metrics, including the Akaike Information Criterion (AIC), negative log-likelihood, mean absolute error, root mean squared error, coefficient of determination (R^2), Continuous Ranked Probability Score (CRPS), and 95% Prediction Interval Coverage Probability (PICP95). For each dataset, four different feature variables are predicted, resulting in a total of 240 GAM models and 240 DGAM models.

The calculation results clearly show that, when evaluated using likelihood-based

and probabilistic performance measures, the Double Generalized Additive Model consistently outperforms the Generalized Additive Model, demonstrating substantially superior predictive performance.

School of Mathematical Sciences
and Geoinformatics
Academic Year 2025

Student's Signature Natkon
Advisor's Signature J. Tanthanudh
Co-Advisor's Signature Amornrat