شبكة المعلومات الجامعية

بسم الله الرحمن الرحيم

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأفلام قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأفلام بعيدا عن الغبار

في درجة حرارة من ١٥-٢٥ مئوية ورطوبة نسبية من ٢٠-٤٠%

To be Kept away from Dust in Dry Cool place of 15-25- c and relative humidity 20-40%

بعض الوثائق الاصلية تالفة

بالرسالة صفحات لم ترد بالاصل

NEW MODELS FOR NONLINEAR REGRESSION AND FOR HANDLING MISSING DATA

By
Mostafa Mohamed Adel Abd-ElKader Hassan
B.Sc. in Computer Engineering – Cairo University

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
in
Computer Engineering

Faculty of Engineering, Cairo University
Giza, Egypt
2007

NEW MODELS FOR NONLINEAR REGRESSION AND FOR HANDLING MISSING DATA

By
Mostafa Mohamed Adel Abd-ElKader Hassan
B.Sc. in Computer Engineering – Cairo University

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
in
Computer Engineering

Supervised by
Prof. Dr. **Amir Fuad Surial Atiya**
Professor, Faculty of Engineering, Cairo University

Dr. **Raafat Said Aly ElFouly**
Assistant Professor, Faculty of Engineering, Cairo University

Faculty of Engineering, Cairo University
Giza, Egypt
2007

NEW MODELS FOR NONLINEAR REGRESSION AND FOR HANDLING MISSING DATA

By
Mostafa Mohamed Adel Abd-ElKader Hassan
B.Sc. in Computer Engineering – Cairo University

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
in
Computer Engineering

Approved by the Examining Committee:

Prof. Dr. **Amir Fuad Surial Atiya**, Thesis Main Advisor

Prof. Dr. **Nevin M. Darwish**, Member

Samia Mashali

Prof. Dr. **Samia A. Mashali**, Member

Faculty of Engineering, Cairo University
Giza, Egypt
2007

Abstract

This work is one step towards better data mining. To be more specific, it addresses the problem of nonlinear regression in case of complete and missing data. We first introduced the Penalized likelihood regression model. This concept vies to achieve a compromise between goodness of fit (as typified by the likelihood function) and smoothness of the data. Attempts to consider the general multidimensional case have been limited. In our work, we propose a new multidimensional penalized likelihood regression method. The approach is based on proposing a roughness term based on the discrepancy between the function values among the K-nearest-neighbors. The proposed formulation yields a simple solution in terms of a system of linear equations. We show that the proposed model is fairly versatile in that it exhibits nice features in handling user defined function constraints and data imperfections. Experimental results confirm that it is competitive with the Gaussian process regression method (one of the best methods out there), and exhibits significant speed advantage.

After that we handle the case of missing data. We developed a method using an ensemble-network model. The proposed method is based on utilizing the inherent uncertainty of the missing records in generating diverse training sets for the ensemble's networks. Specifically we generate the missing

values using their probability distribution function. We repeat this procedure many times thereby creating a number of complete data sets. A network is trained for each of these data sets, thereby obtaining an ensemble of networks. Several variants are proposed, and we show analytically that one of these variants is superior to the conventional mean substitution approach. Simulation results confirm the general superiority of the proposed methods compared to the conventional approaches with respect to nonlinear regression with and without missing data.

Acknowledgments

All thanks and praise go to ALLAH who gave me the strength that enabled me finish this work.

I am really much grateful to my supervisor, Dr. Amir Atiya, who did not spare any effort to help me. He supported me all the way. He was patient and supportive, and provided many invaluable pieces of advice. I also would like to acknowledge my other supervisor, Dr. Raaft El-Fouly, for his cooperation and support.

Many thanks are due to our department chairman, Dr. Nevin Darwish, for her great help, specially with formal departmental issues.

Finally, this work is dedicated to my Mother, Father, and all my family for their complete support and prayers.