Regularization Parameter Tuning Optimization Approach in Logistic Regression

Ahmed El-Koka*, Kyung-Hwan CHA*, Dae-Ki KANG *

*Department of Computer and Information Engineering, DSU (Dongseo University), Korea eng.koka@gmail.com, khcha@gdsu.dongseo.ac.kr, dkkang@dongseo.ac.kr

Abstract— Under regression analysis methods, logistic regression comes and it got popular since it has proved its effectiveness in modelling categorical outcomes as a function of either continuous -real value- or categorical -yes vs. no- variables. The coefficients of this prediction function are based on a data set that is used to shape this function. However, sometimes the dataset, which is used to generate the prediction function of the logistic regression, would have some odds and need to be smoothened to avoid under or over-fitting. Thus, a mathematical regularization part has been introduced to be added to the cost function of logistic regression and it mainly contains an important parameter which is called the regularization parameter that would have to be determined. Often, this regularization parameter is pre-set or pre-expected by the code developer. Few random values would be tested and judged by the accuracy rate of the prediction function applied on the testing set. Obviously, it's neither effective nor practical approach to choose a parameter that would concretely play an important role in varying the accuracy of the prediction function. In this paper, we propose a mathematical approach to tune the regularization parameter in order to achieve the highest accuracy rate possible. Our idea is basically vectorizing the value of the regularization parameter that is fed to the cost function for the purpose of testing its accuracy. Hence, a vectorized value of the accuracy would be produced and each will correspond to a specific value of the regularization parameter value which produced the highest accuracy rate would be the obvious suitable choice for the prediction function.

Keywords — machine learning, logistic regression, regularized logistic regression, regularization parameter tuning.



Ahmed El-Koka is currently a master graduate student at Dongseo University, Pusan, South Korea majoring in Software engineering. He earned his bachelor degree in Electronics Engineering majoring in Robotics and Automation from Multimedia University, Melaka, Malaysia in 2011. Before he joined Dongseo University, He worked at Galactic Advance Engineering (M) Sdn. Bhd. His research interests mainly include Artificial intelligence, Smart services, Game theory, and Complex systems.



Dae-Ki Kang is an assistant professor at Dongseo University in South Korea. He was a senior member of engineering staff at the attached Institute of Electronics & Telecommunications Research Institute in South Korea.

He earned a Ph.D. in computer science from Iowa State University in 2006. Prior to joining Iowa State, he worked at two Bay-area startup companies and at Electronics and Telecommunication Research Institute in South Korea. He received a science master degree in computer science at Sogang University in 1994 and a bachelor of engineering (BE) degree in computer science and engineering at Hanyang University in 1992.

His research interests include Cloud Computing, Big Data, Econometrics, Game Theory, Multi-Relational Learning, Social Networking Services and Complex System Network, Statistical Graphical Model, Ontology Learning, Datamining based Intrusion Detection, Context Sensitive Mobile Advertising, Malicious Bots Detection, Whole Body Interaction, Web application firewall, Web Mining and Computer Vision.



Kyung-Hwan Cha received the B.S. and M.S. and Ph.D. degrees in Electronics Engineering from Pukyung National University, Busan, Korea in 1985, 1990, and 1996, respectively. He was a researcher in LG Research Center from 1990 to 1995. Since 1995, he has been a Professor of the Division of Information and Network Engineering, Dongseo University. His current research interests are in the areas of Digital Signal Processing and Embedded System.