Q2016 Quality Conference course
Theme of the Course/Workshop: Introduction to Big Data and Machine Learning: An application to Official Statistics
Facilitators: Luis F. Lago-Fernández and Gonzalo Martínez-Muñoz, Universidad Autónoma de Madrid

Programme

Time slot	Topic
09:30 – 09:45	Presentation of the lecturers and description of the objectives of the course
Morning session	**Introduction to machine learning and Support Vector Machines**
09:45 – 11:00	Introduction to Machine Learning in the context of Big Data
11:00 – 11:30	Coffee-Break
11:30 – 12:15	Data quality and data auditing
12:15 – 13:00	Support vector machines
13:00 – 14:00	Lunch Break
Afternoon session	**Ensembles of classifiers and an application**
14:00 – 15:30	Decision Trees and Random Forests
15:30 – 16:00	Coffee-Break
16:00 – 17:00	An application of machine learning to official statistics

Description of the course

This course is an introduction to machine learning techniques and to its relation with big data. The course describes the standard machine learning pipeline: data acquisition, auditing, pre-processing and model selection and validation. In the pre-processing stage the concepts of data normalization, standardization or treatment of missing values and outliers are covered. Then, the course focus on two of the most effective learning models nowadays, that are support vector machines (SVM) and random forests. Finally, an application of these methods in the context of official statistics is shown.

Objectives of the course

At the end of this course the participants will be able to identify the different parts of the pipeline of a machine learning application procedure. He/She will be able to apply different strategies to filter missing values and to train SVMs and random forests and validate their results.

Participant's profile

The course is addressed to people working with microdata in the context of official statistics that are interested in extracting further information from the data and to anyone interested in knowing about the different possibilities that machine learning techniques can offer.

Instructor's profile

Gonzalo Martínez-Muñoz received the university degree in Physics (1995) and Ph.D. degree in Computer Science (2006) from the Universidad Autónoma de Madrid (UAM). From 1996 to 2002 he worked in industry. Until 2008 he was an interim assistant professor in the Computer Science Department of the UAM. During 2008/2009, he worked as a Fulbright postdoc researcher at Oregon State University in the group of Professor Thomas G. Dietterich. He is currently a professor at Computer Science Department at UAM. His research interests include machine learning, computer vision, pattern recognition, neural networks, decision trees, and ensemble learning.

Luis F. Lago-Fernández received the B.Sc. degree in Theoretical Physics from the Universidad Autónoma de Madrid, Spain, in 1998, and the Ph.D. degree (cum laude) in Computer Science from the same university in 2003. He is currently Profesor Contratado Doctor in the Computer Science Department, at the Universidad Autónoma de Madrid, Spain, and Scientific Collaborator for the data mining company Cognodata, Madrid, Spain. His research interests include machine learning, data mining and computational neuroscience.