



Contribution ID: 68

Type: not specified

Machine learning for scientific applications

Machine learning for scientific applications

J. Vega

Laboratorio Nacional de Fusión. CIEMAT.

Avenida Complutense, 40. 28040 Madrid (SPAIN)

jesus.vega@ciemat.es

ABSTRACT

At present, a major problem in many scientific fields is not the lack of data but the amount of stored data (that includes waveforms and video-movies). All the data are of no value without mechanisms to efficiently and effectively extract information and knowledge from them. Data mining is the scientific discipline to deal with this. The main distinguishing characteristic of data mining is its “data driven” nature, as opposed to other methods that are often “model driven”. In statistics, researchers frequently try to find the smallest amount of data that gives sufficiently confident estimates. In data mining, we intend to use the opposite approach, namely, we are interested in building a model that is not too complex but at the same time describes the entire database.

Data mining techniques use machine learning. The formulation of a learning problem is rather broad. It encompasses many specific disciplines but this introductory talk will only consider the ones related to pattern recognition (SVM, neural networks, clustering and combination of classifiers) and regression estimation (SVR). Concepts and methodologies will be presented. Also, different ways of estimating the accuracy and reliability of the predictions will be mentioned (Bayes classifiers, logistic regression, conformal predictors, Venn predictors and ridge regression confidence machines). Finally, general ideas about advanced techniques (active learning) and specific applications to nuclear fusion will be shown.

Primary author: Dr VEGA, JESUS (ASOCIACION EURATOM/CIEMAT)