

Models for Data or Models for Prediction?

Gilbert Saporta

CEDRIC-CNAM

Conservatoire National des Arts et Métiers

292 rue Saint Martin

75141 Paris cedex 03, France

Gilbert.Saporta@cnam.fr

Abstract The classical view off statistical modelling consists in establishing a parsimonious representation of a random phenomenon, generally based upon the knowledge of an expert of the application field: the aim of a model is to provide a better understanding of data and of the underlying mechanism which have produced it. On the other hand in Data Mining and Statistical Learning predictive models are merely algorithms and the quality of a model is assessed by its performance for predicting new observations. In this communication, we develop some general considerations about both aspects of modelling.