Preface

The ARTIFICIAL LEARNING is interested in writing computer programs capable to improve itself automatically in time, either on the basis of their own experiment, or starting from former data provided by other programs. In the relatively young scientific field of informatics, the artificial learning plays an increasingly essential part. At the beginning of its existence, in the years 1950, informatics was mainly summarized to program machines by specifying, line after line, the sequence of instructions which the computer would have to follow. Around the years 1990, the software had become so complex that an alternative impose itself naturally: to develop techniques so that the programs can be trained starting from examples. The result is that there exist today many informatics applications in which the methods of the artificial learning are employed to train the software. Better, the resulting code goes far beyond in performance the most led achievements of the manual programmation "line after line". At the time when we pass the first fifty years of informatics to the fifty next ones, it seems certain that the role of the artificial learning will not cease growing in the center of this science.

My purpose in writing this lecture notes has been to give a systematic introduction of major concepts and methodologies of artificial learning from data and to present a unified framework that makes the subject mare accessible to students.

The main audience was the students following the Doctorate courses at the Department of Statistical Sciences, University of Padua and the terminal-year students in informatics at the Faculty of Mathematics, University of Bucharest.

The background material needed to understand these notes is general knowledge of some basic topics in probability and statistics, differential equations and linear algebra, and multivariate calculus.

This course is organized in five chapters. Chapter 1 introduces the reader to the key concepts of machine learning. The general model of learning from examples is presented and the main inductive principles which implement the model are studied. The evaluation of the learning performances and the comparison of the learning methods are discussed in this chapter too.

Learning rules for single perceptron and single-layer nets are covered in Chapter 2. Only supervised rules are considered. The presentation of this learning rules is unified in the sense that they may all viewed as realizing incremental steepest-gradient-descent search on a suitable criterion function.

Chapter 3 deals with learning in multilayer artificial neural nets. It extends the gradient-descent-based learning to multilayer feedforward nets, which results in the back error propagation learning rule (or backprop). An extensive number of methods and heuristics for improving backprop's convergence speed and solution quality are presented, and an attempt is made to give a theoretical basis for such methods and heuristics.

Chapter 4 manly deals with theoretical foundations of the computational capabilities of artificial neural networks. The necessary bounds on the size of the neural networks based multilayer classifier for the cases of training data in general position and in arbitrary position are derived. Theoretical results on continuous function

approximation capabilities of feedforward nets are summarized. The chapter concludes with a discussion of the computational effectiveness of neural net architectures and the efficiency of their hardware implementation.

Chapter 5 present one of the most important applications of the statistical learning: the data mining. The chapter summarizes the principal applications of data mining see as knowledge discovery in data and presents the principal research prototypes and system products. The chapter discusses the social impacts and trends in data mining.

I hope that this lectures notes will prove useful to those students who are interested not only in understanding the underlying theory of machine learning from data but also in pursuing researching this area. A list of relevant references is included with the aim of providing guidance and direction for the reader's own search of the research literature.

Acknowledgement. My deep gratitude goes to Professor Fortunato Pesarin of the Department of Statistical Sciences, University of Padua who invited me to teach this course in the autumn term of the university year 2003-04.

I am grateful also to Professor Luigi Salmaso for the useful discussions and help.

The Author