el código desarrollado, los

datos aplicados, los resultados y unas conclusiones del trabajo

**Universidad Rey Juan Carlos**

**Máster Oficial en Visión Artificial**

**“Biometría Informática”**

**Curso académico 2012/2013**

**Orión García Gallardo**

**Práctica 1: Comparativa de técnicas de clasificación supervisada.**

**Introducción**

There are several applications for Machine Learning (ML), the most significant of which is

predictive data mining. Every instance in any dataset used by machine learning algorithms

is represented using the same set of features. The features may be continuous, categorical binary. If instances are given with known labels (the corresponding correct outputs) then

the learning is called supervised, in contrast to unsupervised learning, where instances are

unlabeled (Jain et al. 1999).

Numerous ML applications involve tasks that can be set up as supervised. In the present

paper, we have concentrated on the techniques necessary to do this. In particular, this work is

concerned with classi?cation problems in which the output of instances admits only discrete,

unordered values. We have limited our references to recent refereed journals, published books

and conferences. A brief review of what ML includes can be found in (Dutton and Conroy

1996). De Mantaras and Armengol (1998) also presented a historical survey of logic and

instance based learning classi?ers.

After a better understanding of the strengths and limitations of each method, the possibility

of integrating two or more algorithms together to solve a problem should be investigated. The

objective is to utilize the strengths of one method to complement the weaknesses of another.

If we are only interested in the best possible classi?cation accuracy, it might be dif?cult or

impossible to ?nd a single classi?er that performs as well as a good ensemble of classi?ers.

Our next section covers wide-ranging issues of supervised machine learning such as data

pre-processing and feature selection. Logic-based learning techniques are described in Sect.

3, whereas perceptron-based techniques are analyzed in Sect. 4. Statistical techniques for ML

are covered in Sect. 5. Section 6 deals with the newest supervised ML technique—Support

Vector Machines (SVMs). In Sect. 7, a representative algorithm for each learning technique

is compared to a number of datasets in order for researchers to have baseline accuracy for new

algorithms in these well-known datasets. Section 8 presents the recent attempt for improving

classi?cation accuracy—ensembles of classi?ers. Finally, the last section concludes this

work.

Descripción de la solución