International Journal of Innovative
Computing, Information and Control
Volume **6**, Number **12**, December **2010**

ICIC International ©2010 ISSN 1349-4198
pp. 5597–5608

ASYMMETRIC SUPPORT VECTOR MACHINE FOR THE CLASSIFICATION PROBLEM WITH ASYMMETRIC COST OF MISCLASSIFICATION

Zhongsheng Hua, Xuemei Zhang and Xiaoyan Xu

School of Management
University of Science and Technology of China
Hefei 230026, P. R. China
{ zshua; xxy204 }@ustc.edu.cn; xmz@mail.ustc.eu.cn

Received September 2009; revised January 2010

Abstract. *Classification algorithms are usually developed under the assumption that costs of different types of misclassification are equal. In many real world classification problems, however, costs of different types of misclassification are asymmetric, and one type of misclassification is more serious than the other. In this paper, we propose an asymmetric support vector machine (ASVM) for the binary classification problem with asymmetric cost of misclassification. Through introducing some adjustable parameters into the model of the conventional SVM, we establish an explicit connection between the expected cost of misclassification (ECM) and the adjustable parameters. The proposed ASVM effectively decreases ECM through minimizing an upper bound of ECM. Computational results verify the efficacy of ASVM compared with the conventional SVM and the method proposed in the previous literature.*
Keywords: Support vector machine, Binary classification, Asymmetric cost of misclassification, Expected cost of misclassification

1. **Introduction.** In a binary classification problem, a training dataset with l samples $(\mathbf{x}_i,\ y_i)$ is given to a decision maker, where the binary outputs $y_i = \pm 1$ correspond to two classes. The two classes labeled –1 and 1 are called negative class and positive class, respectively. For each sample i $(i = 1, 2, \ldots, l)$ in the training dataset, the decision maker observes an input vector $\mathbf{x}_i \in R^n$, and a label y_i indicating one of the two classes to which the sample belongs. The task of the classification problem is to derive from the training dataset a good classifier, so that once $\mathbf{x}$ value of a new subject from a target population is given, a class label of the new subject can be correctly identified by the classifier [1,2]. The binary classification problem described above has been intensively investigated based on different classification rules, e.g., Support Vector Machine (SVM) [3-8], Bayesian rule [9,10], logistic regression [11,12] and neural network [13,14], with the objective of minimizing the expected rate of all types of misclassification. This objective of classification is appropriate when the costs of different types of misclassification are equal.

In many real world classification problems, however, costs of different types of misclassification are asymmetric, and one type of misclassification is more serious than the other. In disease diagnosis problem, for example, a healthy person (in the negative class) may be misdiagnosed to be a patient (in the positive class), which is termed false positive. Similarly, we can define false negative. The cost of false positive is usually limited by excessive medical expenses (including costs of unnecessary medical treatment and further inspections), while the cost of false negative may be an irreparable damage on a patient's health.