

IS MY BANANA RIPE? AN INTRODUCTION TO LEAST SQUARES SUPPORT VECTOR MACHINES

Alan Kimber¹ and Carmen Ybarra Moncada²

¹ School of Mathematics, University of Southampton, Southampton, UK

² Section of Quantitative Biology and Statistics, University of Reading, Reading, UK

This study was motivated by chemometric problems in which NIR spectroscopy methods are used to estimate, for example, the ripeness of a piece of fruit or the nitrogen content of manure. Standard approaches include principal components regression and partial least squares. More recently least squares support vector machines (LS-SVM) have been applied to such “ $p > n$ ” regression problems by scientists in a number of disciplines. In this paper we illustrate the use of LS-SVM on some chemometric data sets. In order to implement the method it is necessary to choose the values of its two hyper-parameters. Poor choices may lead to serious under-fitting or over-fitting. Some consideration is given to methods for choosing appropriate values. In addition, we investigate some statistical properties of LS-SVM, including mean squared prediction error and robustness.