Mobile Phone Sales Forecast Based on Support Vector Machine

Summary

In this paper, factors such as price, wear resistance, resistance to fall, charging interval, battery life, communication stability, photo effect, appearance design, memory size and whether to buy again are taken as input variables and mobile phone sales foreground grade is taken as the output variable. It is observed that the forecasts made by the model built on Support Vector Machine Regression algorithm (SVMR) is consistent with the actual sales of the various mobile phones available in the market.

With rapid strides in the information technology sector of all the countries in the world, the communication industry has become an integral part of any nation’s economy. Many prediction models these days make use of mathematical statistics, grey theory, linear regression, neural networks, genetic algorithms and support vector machines. A majority of these predictions are based on the manufacturer’s sales volume and related information of the past. Support Vector Machine (SVM) was first proposed by Corinna Cortes and Vapnik in 1995. It is a sample data processing model based on statistical theory. It is used to achieve high efficiency in prediction models as it overcomes problems such as local optimum and long convergence time that are faced by BP neural network models. Therefore, SVM is employed to predict the sales of mobile phones and the same is discussed in this paper.

There is a non-linear relationship between the sales of mobile phones and the various factors that influence their sales. Therefore, an SVM can map related data to the high-latitude feature space in a non-linear manner, and thereby constructs a linear regression in the high-latitude feature space, so as to obtain the non-linear regression effect in the low-dimensional space. The optimal plane to be sought is to minimize the "total deviation" of all the sample points from the hyperplane. And also, the presence of sample points between the two boundaries is equivalent to the problem of maximum separation.

The model discussed here composes of three layers. Layer 1 is the input layer for the dependent variable, the layer 2 is the intermediate processing layer and the 3rd and final layer is the output layer. In this model, x1-xn are the low-dimensional input vectors, Xi=(x1, x2, ... xn); K(xi, χ) = <ω(xi) . ω(χ)> is a kernel function, χ_i is a support vector, ω is a function of high-latitude space mapping. The linear combination function f(x) is obtained by concatenating the input vector and the kernel function by the Lagrangian coefficients (β_1, β_2, ... ,β_n) of the corresponding support vectors.

$$f(x) = \operatorname{sgn}[\sum^{n}_{i=1} \beta_i K(x_i,x) + b]$$

The kernel function in the support vector machine regression model mainly includes linear kernel function, polynomial kernel function, sigmoid kernel function, RBF kernel function, etc. Grid search is used to select the optimal parameters.

The predicted value of the model based on SVM is very consistent with the actual sales value announced by each mobile phone manufacturer.