Support Vector Machines

16BCE1259
Shushil Kumar Ravishankar

What are C, Gamma and Kernel parameters of SVM?

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning 'far' and high values meaning 'close'. The gamma parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors.

The C parameter trades off correct classification of training examples against maximization of the decision function's margin. For larger values of C, a smaller margin will be accepted if the decision function is better at classifying all training points correctly. A lower C will encourage a larger margin, therefore a simpler decision function, at the cost of training accuracy.

SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example, ***linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.***

SVR

Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences.

First of all, because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem.

But besides this fact, there is also a more complicated reason, the algorithm is more complicated therefore to be taken in consideration. However, the main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated.

I have selected “Sample Earthquake Data” for regression analysis. This dataset has 22 attributes such as ‘Time’, ’Latitude’, ’Longitude’, ’Depth’, ’Magnitude’, etc. and has 1370 rows.

I have chosen ‘Latitude’ and ‘Longitude’ as parameters to determine the magnitude range (0 if mag<4.5, 1 if mag>4.5)

After splitting the dataset into test and train, we trained the SVR model with the train data.

This model is then used to predict the class of magnitude the earthquake in that particular latitude and longitude lies in.

The accuracy of the model is around 56.4%.

SVM Demo

Observations:

- RBF kernel performs well in all types of dataset.
- Linear kernel performs well only in linearly separable data.
- When gamma is very small, the model is too constrained and cannot capture the complexity or "shape" of the data. The region of influence of any selected support vector would include the whole training set. The resulting model will behave similarly to a linear model with a set of hyperplanes that separate the centers of high density of any pair of two classes.
- 'C' behaves as a regularization parameter in the SVM. A lower C will encourage a larger margin, therefore a simpler decision function, at the cost of training accuracy.