Using Support Vector Machines to classify multidimensional data

Author:
Václav Bartoš
ibartosv@fit.vutbr.cz

Classifying data is a common task in machine learning and other related areas of computer science. It's a problem, where we have a number of points (vectors) in multidimensional space, which are partitioned into two or more classes, and we have to predict, for a new unlabeled point, into which class it belongs.

A linear classifier uses only two classes and needs data to be linearly separable, i.e. there must exist a hyperplane dividing space into two parts – each containing all data points from one category. Classifier then finds such hyperplane and classifies new data point based on the side of hyperplane it appears on.

But there may be more hyperplanes dividing the data. The optimal one is a hyperplane that has the largest distance from both classes (where distance to a class is distance to the nearest point of that class). When such hyperplane is chosen, the classifier is called Support Vector Machine (SVM).

So basic SVM is a method which takes a set of already classified points (training set) and uses simple mathematical formulas and algorithms of quadratic programing to find a hyperplane dividing the training data with the largest possible margin.

There are various extensions to the basic SVM such as Transductive SVM, which improves results by utilizing additional unlabeled data points in the training set, or Soft Margin SVM, which works even if a small portion of training data is wrongly classified.

SVM in its basic form is a linear classifier, but the requirement of linear separability of data points is not satisfied in many cases. A solution is called a "kernel trick". This method is based on transformation of input data into a space of more dimensions in such a way, that they become linearly separable in this space.

In fact, the data is not transformed, but the key part of hyperplane computation – the dot-product function – is replaced by some other, non-linear, function, called kernel function. This is equivalent to finding hyperplane in more-dimensional space. There is a number of kernels used in practice, each suitable for different kind of data. Effectiveness of SVM depends only on kernel and its parameters.

With the kernel trick, SVM can be used as a non-linear classifier. In fact, the kernel trick is often viewed as an integral part of SVM method and usage of SVM implicitly assumes use of some kernel function.

In the presentation, problem of classification will be formally defined, basic method of Support Vector Machines will be described and its theoretical properties and aspects of practical usage will be discussed. Also, some of its extensions will be described. Because the kernel trick is very closely tied with SVM, but it's not easy to understand, a large part of presentation will be focused on its definition and detailed description, including some examples. In the end, a general method how to extend a binary (two-class) classifier into a multi-class one will be described.