Support Vector Machines for Classification

TECHNICAL REPORT CS-RHUL

Jorge Maidana

19th June 2017

0.1 Introduction

Classification is a common task in Machine Learning. Given a dataset with *training* elements, the goal is to find the classes to which new data points, called *test set*, belongs.

Support Vector Machines is build upon the concept of *maximal margin hyperplane*[1]. In a p-dimensional setting, a Support Vector Machines model will draw points of the dataset as $p-$dimensional vectors and a hyperplane will be a flat affine subspace in a $p-1$ dimension with the goal of separate or classify the data points[2].

In such setting, many hyperplanes can be drawn as ilustrated in figure. Since these hyperplanes are based on the training set, we may think on choosing one such that it can be accurate on a test set. An idea of how to do this is to select a *marginal* hyerplane. That is, to calculate the minimal perpendicular distance between the data points and to select the hyperplane which is farthest from these datapoints. Such a hyperplane is known as *maximal margin classifier*.

An important drawback of the marginal classifier is its sensitivity to the points of the dataset. A change in a single $p-vector$ means that a different hyperplane must be needed. SVM is a method that overcome that disadvantage while offering a greater robustness and a better classification for most of the data points. A *support vector classifier* sometimes called a *soft-margin classifier* can separate the data with a wider margin, one which is not as strict as the maximal margin classifier, based on the idea of allow some observations to be missclasified. The observations that lies in the margin or wrong side of the margin are known as *support vectors*. (Figure needed).

Is too pedantic at the moment...

0.2 Support Vector Machines

Support vector classifier are linear classifiers so it expects the to perform well on linearly separable data(Ideas not introduced or connected with before..).

A support vector machines is a generalization of a marginal optimal hyperplane classifier in which the feature space is extended with kernels.(Not introduced).

0.2.1 Derivation

An intuitive idea for looking at Support Vector Machines is to think in terms of search for the "widest strip" between two group of observations as illustrated in().

(Derivation from the course of MIT and LSSVM book)

Say:

Definition of the projection of an uknown vector on the "width-normal" vector.

Trick with +/- samples

Finding the "width" of the strip.

Application of the lagrangian to the width.

Definition of a Quadratic Optimization Problem.

Observing the properties of such definition.

Kernels?

0.3 Future Addendum

0.3.1 Support Vector Classifier

In this example we will fit a dataset randomically generated with the statistical language R

The figures of R are odd and awful.

Bibliography

[1] J.A.K. Suykens, T. Van Gestel, and J. De Brabanter. *Least Squares Support Vector Machines*. World Scientific, 2002.

[2] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer, edition edition, 2015.