Ministry of Education and Science of the Russian Federation
Peter the Great St. Petersburg State Polytechnic University
Institute of Computer Sciences and Technologies
**Graduate School of Cyber-Physical Systems and Control**

# Report, Project

SVM Method for Non-Linear Classification
Discipline: Modern Problems of Informatics and Computer Science
5 May 2017

Student Group: 13541/8

Christopher W. Blake

Professor

Rodionova E.A.

St. Petersburg
2017

# Contents

# Overview

## Support Vector Machines

A support vector machine is a supervised algorithm mainly for classification tasks. A set of previously labeled data is used as training information, and then a line is produced which divides the space into two distinct regions. The idea is similar to regression, but instead of finding a line of best fit, a line of best separation is created. This is demonstrated in figure 1, with traingles on one side of the line and circles on the other.

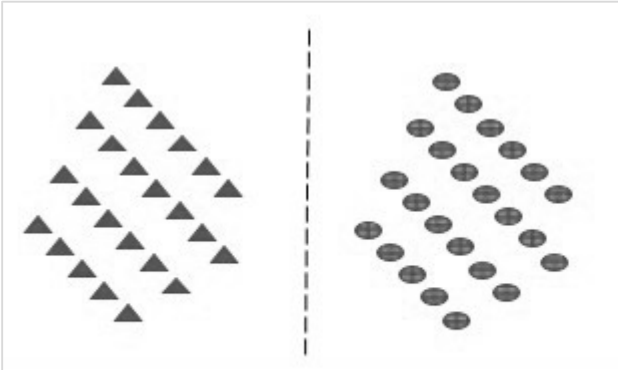

*Figure 1: Linearly Separable Data*

A SVM is inherently a linear classifier, but is extended to non-linear classifications using the kernel technique, which transforms a data set.

## Dimensional Transformation

The classification of nonlinear data is made possible by a kernel function. A kernel function transforms an existing data set into a higher dimensional data set. Below is an example transformation of 2 dimensional data to 3 dimensional data. A simple function of z = x^2 + y^2 is applied, and the data, previously described by X and Y components, is now described by X, Y, and Z components. The data is then explored from the new view possibilities, which is now separable by a line (or hyperplane).

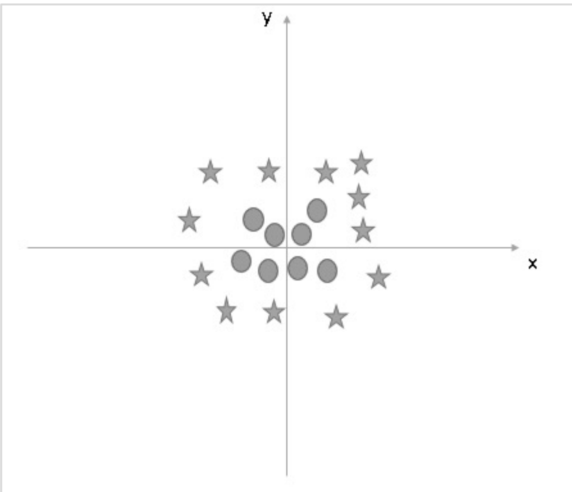


*Figure 2: Non-Convex Data Set, XY Coordinates*

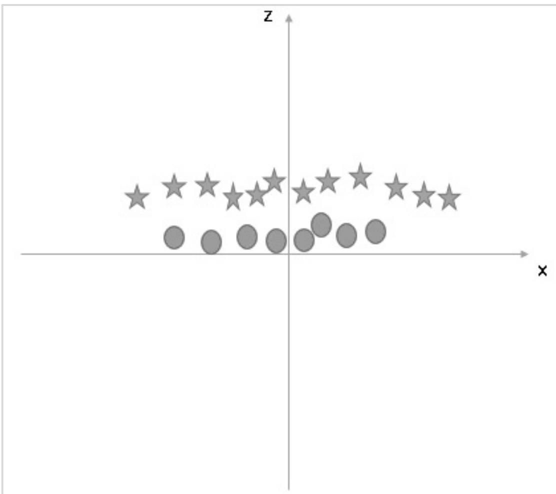


*Figure 3: Transformed Data Set 2D to 3D, XY --> XYZ*

## Example Applications

Below is a brief list of mentioned applications areas where SVM has been utilized.

1. Text and Hypertext categorization
2. Image Classification
3. Image Segmentation
4. Character Recognition (Handwritten)
5. Protein Classification
6. Medical Diagnosis
7. Weather Forecasting

## Separation Line Calculation

The key component to SVM, is the calculation of the separation line. Items laying on one side of the line are a member of the first group, and items laying on the other side are members of the second group.

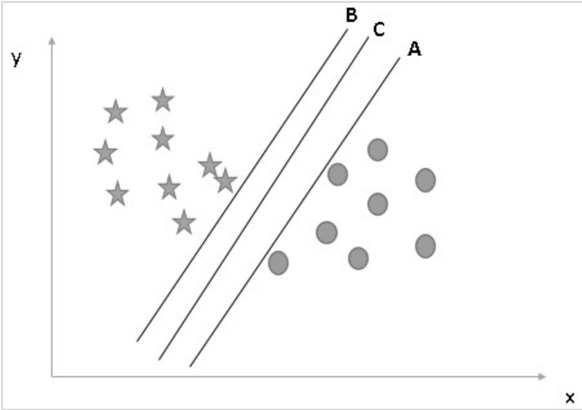

*Figure 4: Determination of Separation Line*

A second part of the problem is determining the best separation line. In figure 4, three lines are shown, each of which successfully separate the data. However, line C creates the largest margin between the two data sets and is clearly preferable, because it is likely more stable.

The process for finding such a separation line is similar to fitting and regression algorithms. A line is chosen and compared to both datasets. Using minimization techniques, the best parameters of the line are chosen which create the largest margin between the two datasets.

### Linear Method

The below method references figure 5, which shows two data sets separate by various lines. The different elements are described below to aid explanations of the method. For the purpose of this example, all data has been normalized to 1.

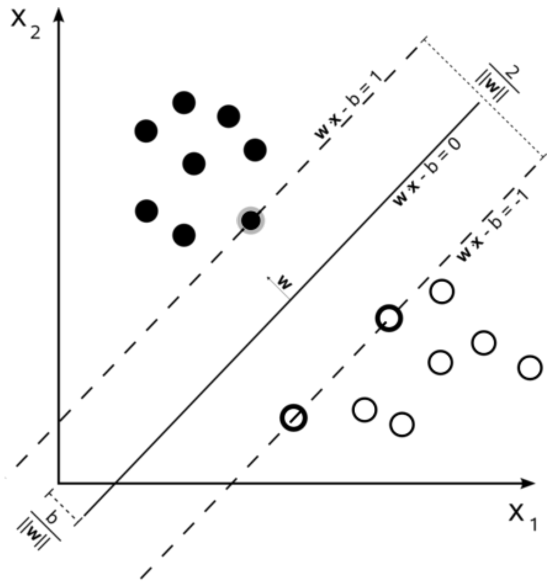

*Figure 5: Datasets separated by maximum margin lines*

X1 – Characteristic 1 of the feature vector
X2 – Characteristic 2 of the feature vector
W – The normal vector of the separation lines
Y – The classification result (1 or -1)

Two parallel lines are selected next to each data set, and the line directly in the middle of them is calculated. The nearest lines are chosen by those points nearest the two data sets, which are called the support vectors.

The initial two lines can be represented as:
Black Dots Line: $\vec{w} \cdot \vec{x_i} - b = 1$
White Dots Line: $\vec{w} \cdot \vec{x_i} - b = -1$

These two equations are combined to calculate the distance between them, which needs to be maximized. This can be done by minimizing the normal vector W.

$$SeperationDistance = \frac{2}{\|\vec{W}\|}$$

The calculation space is further restricted using the previously specified lines separating the black and white does from the margin area.

$$\vec{w} \cdot \vec{x_i} - b \geq 1, if\ y_i = 1\ (black)$$
$$\vec{w} \cdot \vec{x_i} - b \leq -1, if\ y_i = -1\ (white)$$

These restraint equations can be combined as the following.

$$y_i(\vec{w} \cdot \vec{x_i} - b) \geq 1$$

Finally, the minimization of W is combined with the restraints to create the optimization problem. Any multi-dimensional optimization technique may be used at this point.

$$\min\|\vec{W}\|\ \ for\ \ y_i(\vec{w} \cdot \vec{x_i} - b) \geq 1$$

### Non-Linear Extension

The non-linear methods use the dimensional transformation described previously. The dot product of the optimization function is simply replaced with a non-linear transformation function, which is known as kernelling. Below is a list of some common kernels.

- Polynomial (homogeneous): $k(\vec{x_1}, \vec{x_2}) = (\vec{x_1} \,.\, \vec{x_2})^d$
- Polynomial (inhomogeneous): $k(\vec{x_1}, \vec{x_2}) = (\vec{x_1} \,.\, \vec{x_2} + 1)^d$
- Gaussian Radial Basis: $k(\vec{x_1}, \vec{x_2}) = \exp(-\gamma\|\vec{x_1} - \vec{x_2}\|^2)\,, for\ \gamma > 0$
- Hyperbolic Tangent: $k(\vec{x_1}, \vec{x_2}) = \tanh(a\vec{x_1} \cdot \vec{x_2} + c)$
  $for\ some\ a > 0\ and\ c > 0$ (not all)

## Advantages/Disadvantages

| Advantages | Disadvantages |
|---|---|
| 1. Kernel method introduces flexibility.<br>2. Transformations are robust.<br>3. Good out-of-sample generalization.<br>4. Provides a unique solution, unlike neural networks. | 1. Lack of transparency.<br>2. Model selection. |

## History

Below is a list of some of the most notable events relating SVM development to pattern recognition history.

- 1936 R.A. Fisher suggest the first algorithm for pattern recognition.
- 1950 Introduction of the “Theory of Reproducing Kernels”.
- 1957 Frank Rosenblatt introduces the perceptron.
- 1963 Vapnik and Lerner introduce the Generalized Portrait algorithm.
- 1964 Aizerman, Braverman, and Rozonoer introduce kernels as inner products in a feature space, and their geometrical interpretation.
- 1973 Duda and Hart discuss large margin hyperplanes.
- 1989 Several statistical mechanics papers are using large margin hyperplanes.
- 1990 Poggio, Griosi, and Wahba discuss the use of kernels.
- 1995 Cortes and Vapnik introduce latest iteration known as Soft margin.

## References

1. **Corinna Cortes**. (n.d.) In Wikipedia. Retrieved April 12, 2017, from https://en.wikipedia.org/wiki/Corinna_Cortes
2. **Kernel method**. (n.d.) In Wikipedia. Retrieved April 12, 2017, from https://en.wikipedia.org/wiki/Kernel_method
3. **Positive-definite kernel**. (n.d.) In Wikipedia. Retrieved April 12, 2017, from https://en.wikipedia.org/wiki/Positive-definite_kernel
4. **Support vector machine**. (n.d.) In Wikipedia. Retrieved April 12, 2017, from https://en.wikipedia.org/wiki/Support_vector_machine
5. **Support Vector Machine.** Retrieved April 12, 2017, from https://www.dezyre.com/data-science-in-r-programming-tutorial/support-vector-machine-tutorial
6. Noel Bambrick, A. (2016, July). **Support Vector Machines: A Simple Explanation.** Retrieved April 12, 2017, from http://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html
7. Auria, Laura. Moro, Rouslan. (2008, August). **Support Vector Machines as a Technique for Solvency Analysis.** Deutsches Institut für Wirtschaftsforschung.
8. **Support Vector Machines: History.** Retrieved April 12, 2017, from http://www.svms.org/history.html
9. Ray, S. (2015, October 6). **Understanding Support Vector Machine algorithm from examples (along with code)**. Retrieved April 12, 2017, from https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/
10. **Vladimir Vapnik**. (n.d.) In Wikipedia. Retrieved April 12, 2017, from https://en.wikipedia.org/wiki/Vladimir_Vapnik