Artificial Intelligence & Machine Learning

Artificial Intelligence , Soft Computing, Machine Learning, Computational Intelligence

Support Vector Machines (SVM) Fundamentals Part-II

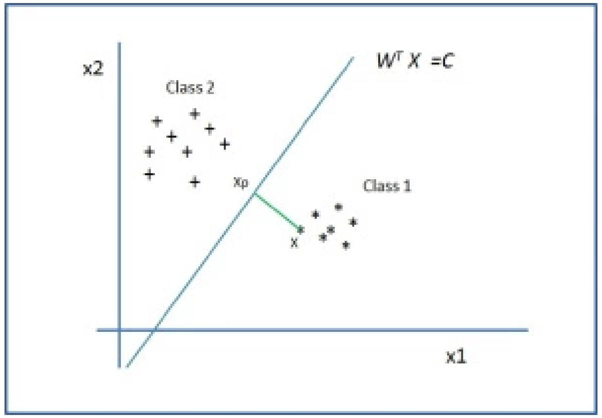
Posted on July 30, 2013October 19, 2013 by panthimanshu17
Hi,

In the last post Support Vector Machines (SVM) Fundamentals Part-I (https://panthimanshu17.wordpress.com/2013/07/28/svm-fundamentals-part-1/) , I talked about basic building blocks of support vector machines (SVM), lets just recap some important points:

1. equation of n-dimensional hyperplane can be written as $W^T X = C$, *where* $W=[w1,w2,\ldots\ldots wn]$ *and* $X=[X1,X2,\ldots\ldots Xn]$
2. hyperplane separates the space into two half spaces (positive half space and negative half space).
3. A hyperplane is also known as linear discriminant as it linearly divides the space in two halfs.
4. Support vector machine is a linear discriminant.

Now lets move ahead. As I have mentioned equation of a hyperplane can be written as $W^T X = C$, we will discuss about some properties W and C

- C determines the position of hyperplane and W determines the orientation (angle with axis) of a hyperplane. How? lets analyze it by taking a 2-dimensional surface, for a two dimensional surface hyperplane will be a line and the equation will be $w1x1+w2x2=C$, this equation can be written as $x2=-(w1/w2)*x1+C$, now compare it with general line equation $y=mx+C$, as you can see m determines the angle with the axis (-w1/w2) and C determines position in the X-Y Plane.
- Vector W is orthogonal (https://en.wikipedia.org/wiki/Orthogonality) to the hyperplane, the direction of W is in the direction of positive half. How? lets take two points m and n in hyperplane, now these points are in hyperplane so they will satisfy the equation $W^T m=C$ *and* $W^T n=C$. Subtracting these two equations we will get $W^T(m-n)=0$, now direction of vector $(m-n)$ will be in the direction of plane and as the dot product of W with $(m-n)$ is zero, it means vector W is orthogonal (perpendicular in layman term) to hyperplane.
- Shortest distance between a point and the hyperplane. For Illustration lets take an example of figure below:



(https://panthimanshu17.files.wordpress.com/2013/07/mindistancefrompoint.jpg)

To find the minimum distance between Point X and and decision boundary, we need to find point Xp in decision boundary such that the vector (Xp-X) is orthogonal to the boundary. This becomes an optimization problem with a objective function:

find Xp Such that ||Xp-X|| is minimum and W^T *Xp =C (as Xp is on decision boundary)*

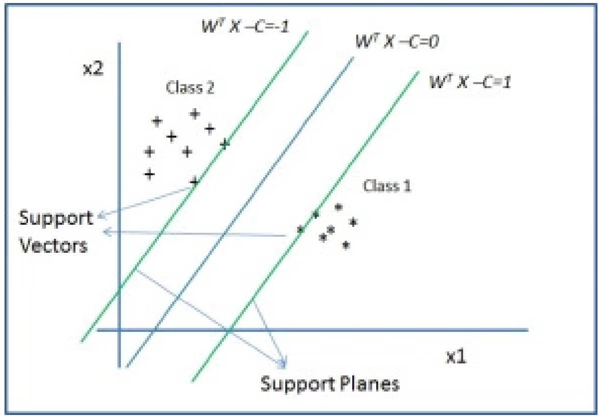
Solving above optimization problem requires Formulation of Lagrangian (http://en.wikipedia.org/wiki/Lagrangian) and applying Karush-Kuhn-Tucker (KKT) (https://en.wikipedia.org/wiki/Karush%E2%80%93Kuhn%E2%80%93Tucker_conditions)conditions. For the sake of simplicity I am just producing Optimization Result. (If you want mathematics involved just send me a mail).

$Xp = X - ((W^T X - C)W/||W||^2)$ and the distance D between Xp and X:

D= $(W^T X - C)/||W||$

The above Distance equation is very important as it forms basis of Support Vector Machines (SVM)

Now I think that we have discussed enough fundamentals and building block so lets dive straight into Support Vector Machines. As I discussed in the last post that support Vector machine is a linear classifier that maximizes the margin. For Illustration purpose lets have a look at figure below:



[(https://panthimanshu17.files.wordpress.com/2013/07/svm.jpg)](https://panthimanshu17.files.wordpress.com/2013/07/svm.jpg)

Support Vector Machine with Support planes and support vectors

Take an example of two class problem classified by hyperplane $W^T X - C=0$, positive class is represented by hyperplane $W^T X - C=1$ and negative class s represented by hyperplane $W^T X - C=-1$, the patterns near the boundaries of each class falls on these hyperplanes. these hyperplanes shown in green in above figure are known as **support planes.** Pattern Vectors lying on the support plane are known as **Support Vectors.** These support vectors are sufficient for training of support vector machine it also results in data reduction.

As Support Vector Machine works on principle of margin maximization. from the above equation distance D between point (support vector) on support plane and and the classifier hyperplane $W^T X - C=0$ is

$$D= (W^T X - C)/||W||$$

now for negative class $W^T X - C=-1$ *and for positive class* $W^T X - C=1$ *so the total margin between two support planes*

$$Dt=2/||W||$$

Support Vector Machine aims to maximize Dt. so it becomes an optimization problem with a statement

$$\text{Maximize } Dt=2/||W||$$

It can be converted into dual problem: Minimize $||W||/2$ *or* $||W||^2/2$ *Subjected to:*

$W^T X - C>=1$ *for positive half patterns and* $W^T X - C<= -1$ *for negative half patterns*

Lets conclude second post here. In the next post (https://panthimanshu17.wordpress.com/2013/08/21/support-vector-machines-svm-fundamentals-part-iii/ (https://panthimanshu17.wordpress.com/2013/08/21/support-vector-machines-svm-fundamentals-part-iii/)) we will look into some more aspects of Support Vector Machines.

my email address is: panthimanshu17@gmail.com

My LinkedIn profile is in about page.

Tagged Artificial Intelligence, Dimension, Hyperplane, Mathematical optimization, support planes, support vector machine, support vector machines, support vectors, SVM

---

6 thoughts on “Support Vector Machines (SVM) Fundamentals Part-II”

1. Pingback: Support Vector Machines (SVM) Fundamentals Part-I | Artificial Intelligence & Machine Learning
2. Pingback: Support Vector Machines (SVM) Fundamentals Part-III | Artificial Intelligence & Machine Learning
3. ***Shreyas*** says: October 19, 2013 at 11:54 am
   Why Do i need to convert Maximation problems to minimization ?

   Reply

   panthimanshu17 says: October 19, 2013 at 12:14 pm
   as ||W||-> 0 2/||W||-> infinity also 2/||W|| is not differentiable at ||W||=0. also aim is to convert objective fun to such a form that it can be solve using standard optimization techniques (QP etc).

   Reply
4. ***Morgan*** says: July 24, 2017 at 1:16 am
   how do one plot the points for the training data, from most of the post on svm I have been seeing, they just explain using a graph with points already plotted and classified and then they explain. But how are those points gotten. What is the formula to plot the points?

   Reply

   panthimanshu17 says: July 24, 2017 at 2:56 pm

For classification problems you generally have training data and labels associated with training data. Based on this you try to fit a model. Data may have many features. In order to visualise data in compressed feature space(2-3), there are many methods available, simplest of them is PCA. You can use PCA to project data into 2 or 3 dim space to visualise or plot.

Reply