Support Vector Machine

Compiled by Koushika B, COE17B044

Guided by

Dr Umarani Jayaraman

Department of Computer Science and Engineering

Indian Institute of Information Technology Design and Manufacturing

Kancheepuram

May 5, 2022

SVM

- $w^t X_i + w_0 > 0$
- $w.X_i + w_0 > 0$ if $Xi \epsilon w_1$
- $w.X_i + w_0 < 0$ if $Xi \epsilon w_2$
- $g(X) = w.X_i + w_0 \geq b$ for good generalization
- The distance of point 'X' from the hyperplane 'H' which is represented as g(X) can be calculated as $r = g(x)/||w||$
- which is nothing but $(W.X_i + w_0)/||w||$.

So, we must ensure that

$$\boxed{(W.X_i + w_0)/||W|| \geq b} \Rightarrow 1$$

$$(W.X_i + w_0) \geq b.||W||$$

$$\boxed{W.X_i + w_0 \geq 1 \text{ if } X_i \epsilon \omega_1}$$

$$\boxed{W.X_i + w_0 \leq -1 \text{ if } X_i \epsilon \omega_2} \Rightarrow 2$$

From 2, we can establish an uniform criteria as follows Y_i is the class label whose values is ± 1

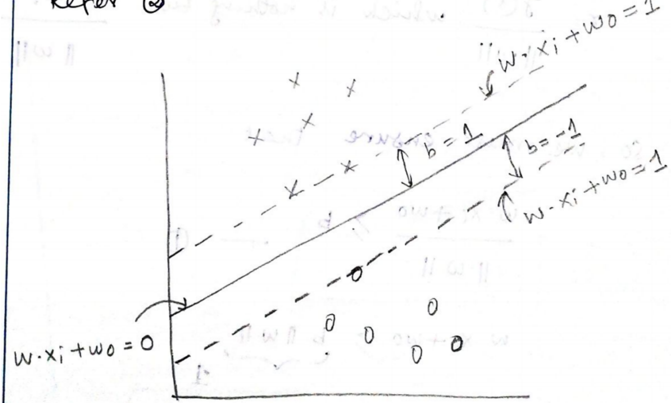
- $+1$ for class ω_1
- -1 for class ω_2

$y_i(W.X_i + w_0) \geq 1$ and this equality holds

$y_i(W.X_i + w_0) = 1$ if X_i are support vectors

$y_i(W.X_i + w_0) > 1$ if X_i are not support vectors

Refer ②



From 1 ,

- In order to maximize the margin 'b', $||w||$ has to be minimized and at the same time w_o has to be maximized .
- For minimization of $||w||$, let's consider the other constraints.

$$\boxed{y_i(w.X_i + w_0) = 1}$$

Because it is constraint optimisation problem, it can be converted into un-constraint problem by using the lagrangian multiplier.

Minimization of $||W||$ is same as minimization of $\phi(W)$ - function of W

$\phi(W) = W^t.W$
$\phi(w) = \frac{1}{2}.W.W$ (dot product)

1/2 is introduced for mathematical convenience.

with the constraints,

$$\boxed{y_i(W.X_i + w_0) = 1}$$

subject to the constraint, if X_i are support vectors

It can be written using unconstraint optimization problem as follows

$$L(||W||, w_o) = \frac{1}{2}||W||^2 - \sum_{i=1}^{n} \lambda[y_i(W.X_i + w_o) - 1]$$

- minimize $||w||$.
- $\frac{1}{2}||W||^2$ is objective function
- $[y_i(W.X_i + w_o) - 1]$ is a constraint

We can define the lagrangian of the form,

$$L(w, w_o) = \frac{1}{2}(w.w) – \sum_{\alpha_i}[yi(w.Xi + wo) – 1]$$

- minimize $||w_o||$.
- maximize $||w||$.
- α_i is Lagrangian multiplier

$$L(w, w_o) = \frac{1}{2}(w.w) – \sum_{\alpha_i}[y_i(w.X_i + w_o) – 1]$$

by taking derivative w.r.t 'w' and $'w_o'$

$$L(W, w_o) = \frac{1}{2}(W.W) - \sum \alpha_i [y_i(WX_i + w_o) - 1]$$
$$L(W, w_o) = \frac{1}{2}(W.W) - \sum \alpha_i y_i (WX_i) - \sum \alpha_i y_i w_o + \sum \alpha_i$$

$$\frac{\partial L}{\partial w_o} = -\sum \alpha_i y_i = 0$$

$$\boxed{\frac{\partial L}{\partial w_o} = \sum_{i=1}^{n} \alpha_i y_i = 0} \Leftarrow \text{one of the constraints}$$

n= number of samples during training

$L(W, w_o) = \frac{1}{2}(W.W) - \sum \alpha_i y_i (W.X_i) - \sum \alpha_i y_i w_o + \sum \alpha_i$
$\frac{\partial L}{\partial w} = W - \sum \alpha_i y_i X_i = 0$

$$\boxed{W = \sum_{i=1}^{n} \alpha_i y_i X_i}$$

Substitute those values in 1
$L(W, w_o) = \frac{1}{2}(W.W) - \sum \alpha_i y_i (W.X_i) - \sum \alpha_i y_i w_o + \sum \alpha_i$

where
$W.X_i = \alpha_j y_j X_j$ and $\sum \alpha_i y_i w_o) = 0$

$$= \frac{1}{2}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j (X_i.X_j) - \sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j (X_i.X_j) + \sum \alpha_i$$

$$L(W, w_o) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j (X_i.X_j)$$

We have to maximize this with different values of αi

α_i - lagrangian multipliers are always positive

$$\boxed{\alpha_i \geq 0}$$

$$\boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$$

- When we try to maximize this lagrangian function, it is quite likely some of the lagrangian multipliers are zero and few of the lagrangian multipliers are very high.
- if $\alpha_i = 0$; that indicates its corresponding training vectors (X_i) are not support vectors.
- if $\alpha_i \neq 0$, its corresponding training vectors (X_i) are having high influence over the position of hyper plane (support vectors).
- Here $\alpha_i \neq 0$ will go into decision making.

$g(z) = W.Z + w_o$
$g(z) = sign(\sum_{i=1}^{n} \alpha_i y_i X_i.Z + w_o)$

unknown Feature Vector Z

- if sign is $+$ve $Z \epsilon \omega_1$
- if sign is -ve $Z \epsilon \omega_2$

The steps of SVM design to estimate W and w_o

$$\boxed{W = \sum_{i=1}^{n} \alpha_i y_i X_i}$$

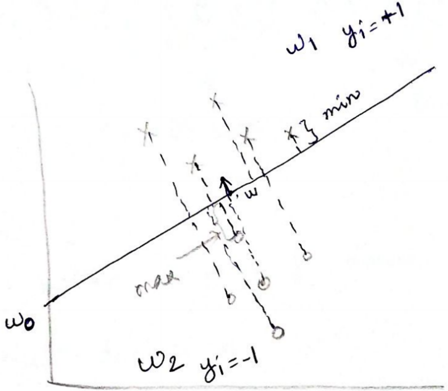
and w_o

$$wo = \frac{1}{2}[min \sum_{\forall \epsilon y_i = +1} \alpha_i y_i (X_i . X_j) + max \sum_{\forall \epsilon y_i = -1} \alpha_i y_i (X_i . X_j)]$$

$$wo = \frac{1}{2}[min \sum_{\forall \epsilon y_i = +1} W . X_i + max \sum_{\forall \epsilon y_i = -1} W . X_i]$$

substitute these W and w_o into classification rule to classify the unknown feature vector Z.

$$g(Z) = W . Z + w_o$$
$$g(Z) = sign \sum_{i=1}^{n} \alpha_i y_i X_i . Z + w_o$$



ω_1 $y_i = +1$
min
w
max
w_0
ω_2 $y_i = -1$

but in implementation,
$w^t X_i + w_o = 1$
$w.X_i + w_o = 1$

$$\boxed{w_o = 1 - (w.X)}$$

Additional slides

Consider the optimization problem maximize f(x,y) subject to :

g(x,y) $=$ 0 or g(x,y) $=$ c
$L(x, y, \lambda) = f(x, y) – \lambda g(x, y)$

f(x,y) is the objective function
g(x,y) is the constraint

assume both f and g have continuous first partial derivatives.

Additional slides

Here

$$\boxed{f(x,y) = f(||w||, w_o)} \Rightarrow \text{objective function}$$

$$\boxed{g(x,y) = y_i(w.X + w_o) = 1} \Rightarrow \text{constraint}$$

we can define the lagrangian of the form ,

$L(||w||, w_o) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \lambda[y_i(w.X_i + w_o) - 1]$

- minimize $||w||$.
- maximize $||w_o||$
- $||w||^2$ is objective function
- $[y_i(w.X_i + w_o) - 1]$ is the constraint

THANK YOU