Zaki (2020) CH21 

# Support Vector Machines



## 5.1 Support Vectors & Margins

D = Dataset w/ n points $x_i$
in d-demenoinl space,
with class labels $y_i \in \{+1, -1\}$

### Hyperplances

A hyperplane in d-dimensions is given as a set of all points $x \in R^{p \text{ (cols)}}$ that satisfy the equation $h(x) = 0$

where $h(x)$ is a hyperplance function

$$h(x) = w^T x + b = w_1 x_1 + \dots + w_d x_d + b$$

$w$ = weight vector, $b$ = scalar bias

Points that create the hyperplane =

$$h(x) = w^T x + b = 0$$

Also defined as set of points that:

$$w^T x = -b$$

recall, the weight vector is orthogonal to the hyperplane $h(x) = 0$

- orthog means perpendicular

$-\frac{b}{w_1}$ = the offset where the hyper intersects the $i$th Dimension ($w_i \neq 0$)

weight vector specifies the direction that is normal to the hyperP

b fixes the offset to the hP is d-space

# Seperating Hyperplane

- A hp spits D-space into 2 halfspaces
- A Dataset is said to be linearly sep is ea Half only containing one class
- if lin sep $= h(x) = 0 : y_1 = 1 = h(x) < 0$ & $y_2 = 0 > 0$
- in this cause the hP equation itself is a linear ~~clarff~~ classifier or linear discriminant
- That is to say it predicts class

$$y = \begin{cases} +1 & \text{if } h(x) > 0 \\ -1 & \text{if } h(x) < 0 \end{cases}$$

## Distance d a point to the HP

Consider a data point $X \in R^d$ which does not sit on the hyperplane

let $X_p$ be the point X orthogonal projection onto the hyperplane (Point on the hp)

Let this distance be r

$r = X - X_p$ (the line)

this means:

$X = X_p + r$ (transform $X_p$ to X w/ r)

$X = X_p + r \frac{w}{\|w\|}$

r = (lowercase) is the Directed distance of the point $X$ to $X_p$

r gives the offset of X from $X_p$ in terms of unit weight vect

$\frac{w}{\|w\|}$ = unit weight vector

r = Posit is r is in same dir as w

Neg y r is in opposite dir to w

Explained

Example:

HP 2-D = $x - 2y + 3 = 0$

Point = $(4, 2)$

goal = Express $X = (4,2)$ as

$$X = XP + r\frac{w}{\|w\|}$$

where XP = Projection of X onto the line

w = weight vector (ORTH TO HP)

r = signed distance (+/-)

Calculations

- $w = (1, -2)$ (COEFF of HP)
- Projection FROM X onto the Hyp = Xp to find = point on line that is closest to X. Done using vector proj or solving linear system of equations
- $Xp = (2, 2.25)$
- Signed Distance (r) calced as dot prod of $(X - Xp) \cdot \frac{w}{\|w\|}$ $\|w\| = \sqrt{(1^2 + -2^2)}$ $= \sqrt{5} \approx 2.2$

$r = X \Rightarrow X_p$ = vector

Bact to textbook: one unit of orth

$$X = X_p + r$$

$$X = X_p + \left(r \frac{w}{\|w\|}\right)$$

Scalar of signed dist

- This is not a replace/Plugin for $r$
- instead it is meant to express $X$ in terms of its projection & distances

How to calculate distance r?

① $X = X_p + r \cdot \frac{w}{\|w\|}$

② sub $X_p = X - X_p = r \cdot \frac{w}{\|w\|}$

③ Dot Prod both sides by $\frac{w}{\|w\|}$

$(x - x_p) \cdot \frac{w}{\|w\|} = r * \left(\frac{w}{\|w\|} \cdot \frac{w}{\|w\|}\right)$

④ Since $\frac{w}{\|w\|}$ = unit vector w/ Mag = 1 then $= \underline{1}$

⑤ final = $r = (x - x_p) \cdot \left(\frac{w}{\|w\|}\right)$

note: w vector always points positive hence the need for a Direction Distance to cover the neighborhood Space class.

The reason we present x as a eqn of $x_p + r\frac{w}{\|w\|}$ is because we now plug into is the hyper-plane func

$$HP = h(x) = w^T x + b = w_1 x_1 + w_2 x_2 + \dots + w_d x_d + b$$

$$h(x) = h\left(x_p + r\, w/\|w\|\right)$$

$$= w^T\left(x_p + w/\|w\|\right) + b$$

$$= \underbrace{w^T x_p + b}_{h(x_p)} + r\, w^T w/\|w\|$$

$$= \underbrace{h(x_p)}_{0} + r\|w\|$$

$$h(x) = r\|w\|$$

Thus, the directed distance of a point to HP=

$$r = \frac{h(x)}{\|w\|}$$

To obtain Distance (no direction & thus non-neg) we multiply r but class label $\{h(x)<0 = -1,\ h(x)>0 = +1\}$

$h(x)=0$

the distance of a point to the HP:

$$\delta = y \cdot r = \frac{y \cdot h(x)}{\|w\|}$$

- for origin the directed Distance to the hyperplane =

$$r = \frac{h(0)}{\|w\|} = \frac{w^T 0 + b}{\|w\|} = \frac{b}{\|w\|}$$

Example:

2-dimensional, hyperplane = line, $x = (x_1, x_2)^T$

$h(x) = w^T x + b = w_1 x_1 + w_2 x_2 + b = 0$

Rearrange to isolate $x_2$:

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{b}{w_2}$$

Slope-intercept

- Why isolate $w_2$? $w_2$ = X-cordinate. Re-arrange for $y = mx + b$ which is a typical representation of line, i.e. y ~~given~~ relationship to x

Here shows
- $\frac{w_1}{w_2}$ to be the slope &
- $\frac{b}{w_2}$ to be the intercept along 2nd Dimension (y)

Consider two points on the HP:

$p = (p_1, p_2) = (4, 0)$ $q = (q_1, q_2) = (2, 5)$

$$-\frac{w_1}{w_2} = \frac{q_2 - p_2}{q_1 - p_1} = \frac{5 - 0}{2 - 4} = -\frac{5}{2}$$

- rise / run

So $w_1 = 5$ & $w_2 = 2$ on the HP

$$h(x) = 5x_1 + 2x_2 + b$$

Can derive b by using any point on the HP

$x = 4, 0$ $\quad b = -5x_1 - 2x_2$

$b = -5 \cdot 4 - 2 \cdot 0$

$b = -20 - 0$

$b = -20$

thus weight vector = $\underset{w}{\begin{pmatrix} 5 \\ 2 \end{pmatrix}}$ $\quad \underset{b}{\text{bias}} = -20$

$$h(x) = w^T x + b = \begin{pmatrix} 5 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 20 = 0$$

Distance d origin from HP =

$$\delta = y \cdot r = \frac{y \cdot h(x)}{\|w\|}$$

$$= -1 \cdot r = \frac{y - b}{\|w\|} = \frac{-(-20)}{\sqrt{29}} = 3.71$$

# Margin & Support Vectors of a HP

given train set D of n points $x_i$ with $y_i \in (+1, -1)$ and hyperplane $h(x) = 0$

For each data point $x_i$ we can find its distance to the hyperplane using:

$$\delta_i = \frac{y_i \cdot h(x_i)}{\|w\|} = \frac{y_i (w^T x_i) + b}{\|w\|}$$

Over all of the points, the margin is the min distance point from HP

$$\delta^* = \min_{x_i} \left\{ \frac{y_i (w^T x_i) + b}{\|w\|} \right\}$$

$\delta \neq 0$ as is assumed to be a seperating HP

- All points that acheive min DST = Support vectors
- Support vector = on margin

Note:

- $y_i (w^T x_i) + b$ = absolute Distance
- $\|w\|$ = makes it relative to unit vector

# Canonical Hyperplane

- Consider the hyperplane equation:
- Multiplying on both sides by some scalar $s$ yeilds an equivalent HP:

$$s \cdot h(x) = s \cdot w^T x + s \cdot b = (sw)^T x + (sb) = 0$$

$sw = 0$

- Simply scaling will not change the HP
- Need to set the target $sb = 1$

- This creates a canonical hyperplane
- Not nessesarily a new HP. It is a HP where the margin function output $= 1$

$$(s \cdot y^*) \cdot (w^T x^* + b) = 1$$

- $y^*$ = margin data point
- need to find the scale that allows $= 1$

implies:

$$s = \frac{1}{y^*(w^T x^* + b)} = \frac{1}{y^* h(x^*)}$$

for the canonical hyperplane, for each support vector $x_i^*$ (label $y_i^*$) we now have $y_i^* h(x_i) = 1$

And hence, any point that is not a support vector is $\geq 1$

All points:

- $y_i (w^T x_i + b) \geq 1$ for all $x_i \in D$

# Canonical hP: Example

Separating Hyperplane: $h(x) = \begin{pmatrix} 5 \\ 2 \end{pmatrix}^T x - 20 = 0$

Support vector $= x^* = (2,2)^T \quad y^* = -1$

To find the canonical hyperplane eq we need rescale the weight vector and bias by the scalar s

$$s = \frac{1}{y^* h(x^*)} = -1 \begin{pmatrix} 5 \\ 2 \end{pmatrix}^T \begin{pmatrix} 2 \\ 2 \end{pmatrix} - 20 = \frac{1}{6}$$

$w \quad x^* \quad b$

thus rescaled w vector = $w = \frac{1}{6} * \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 2/6 \end{pmatrix}$

bias = $b = \frac{-20}{6}$

Canonical form of hyperplane:

$$h(x) = \begin{pmatrix} 5/6 \\ 2/6 \end{pmatrix}^T x - (20/6) = \begin{pmatrix} 0.833 \\ 0.333 \end{pmatrix}^T x - 3.33$$

Margin:

$$\delta^* = \frac{y^* h(x)}{\|w\|} = \frac{1}{\sqrt{\frac{5^2}{6} + \frac{2^2}{6}}} = 1.14$$

Absolute margin

relative margin in terms of w