Support Vector Machine

- To find an optimal seperating hyperplane which maximizes the margin of the training data.
- A supervised learning model, ie have labeled data.
- Used for classification & regression.
- Learns a linear model (hyperplane)

$\boxed{\text{Hyperplane}}$: A generalization of a plane $\boxed{\vec{\beta}^T\vec{x} = 0}$

1 dim (point), 2-dim (line), 3-dim (plane)

Objective To find an optimum ~~operating~~ seperating hyperplane

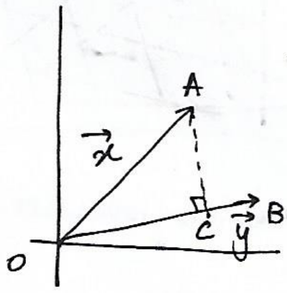
$\boxed{\text{Margin}}$: Twice the distance between a given hyperplane and the closest point.

$\boxed{\text{optimum seperating hyperplane}}$ will be the one with biggest margin, so that It correctly classifies the training data & generalizes the unseen data better.

— X —

$\boxed{\text{BASIC-I}}$

Orthogonal projection of $\vec{x}$ onto $\vec{y}$



$\vec{OA} = \vec{x}$, $\vec{OB} = \vec{y}$ and $\vec{OC} = \vec{z}$

$$\cos\theta = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\,\|\vec{y}\|} \quad \text{(dot product)}$$

Let $\vec{u} = \dfrac{\vec{y}}{\|\vec{y}\|}$ (unit vector in direction of vector $\vec{y}$)

Also $\cos\theta = \dfrac{\|\vec{z}\|}{\|\vec{x}\|} \Rightarrow \|\vec{z}\| = \|\vec{x}\|\cos\theta = \dfrac{\vec{x}\cdot\vec{y}}{\|\vec{y}\|} = \vec{x}\cdot\vec{u}$

Since $\vec{z}$ is in the same direction as $\vec{y}$.

& $\vec{u}$ is the unit vector ie $\vec{u} = \dfrac{\vec{z}}{\|\vec{z}\|} \Rightarrow \vec{z} = \|\vec{z}\|\,\vec{u}$

The vector $\boxed{\vec{z} = (\vec{u}\cdot\vec{x})\vec{u}}$ is the projection of $\vec{x}$ onto $\vec{y}$

Orthogonal vector : Two vectors are orthogonal if there dot product is zero.

Hyperplane $\vec{\beta}^T \vec{x} = \vec{0}$ [Affine subspace of dimension p-1 in p-dimensional space

Consider a 2-dim, hyperplane is a line, say $y = ax + b$

$$\Rightarrow y - ax - b = 0 \Rightarrow \vec{\beta} = \begin{pmatrix} -b \\ -a \\ 1 \end{pmatrix} \quad \vec{x} = \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$$

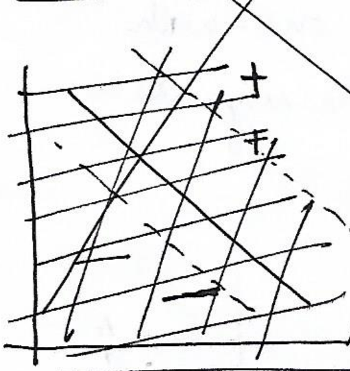
Thus $\vec{\beta}^T \vec{x} = y - ax - b$

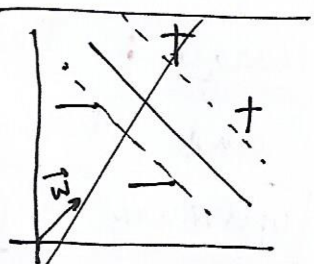
~~Vector~~ $\vec{\beta}$ is ~~nod~~ normal to hyperplane $\vec{\beta}^T \vec{x} = 0$

By definition a hyperplane is defined, we suppose that we have a vector that is orthogonal to the hyperplane.

Normal : A line or vector $\perp$er to a given object.

Decision boundary





MAXIM MARGIN CLASSIFIER

~~Decision boundary~~

Consider a two dimensional space, hyperplane is a line.

say $\boxed{\beta_0 + \beta_1 X_2 + \beta_2 X_2 = 0}$ —(1) for parameters β_0, β_1 & β_2

Eq^n (1) defines the hyperplane means that for any point $X = (X_1, X_2)^T$ lies on hyperplane eq^n (1) holds.

In p-dimensional setting, hyperplane is

$$\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p = 0 \quad -(2)$$

$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$ and $X = (X, X_2 \ldots X_p)^T$, then hyperplane is given by $\beta^T X = 0$

If x doesnot satisfy ② $\Rightarrow \beta^T X > 0$ or $\beta^T X < 0$

Thus, hyperplane can be ~~becomes~~ ~~thoughout of as a~~ thought of as a subspace dividing a p-dim space in 2 halves

Classification using hyperplane

Suppose we have a $n \times p$ data matrix X that consists of n training observation in a p-dimensional space

$$x_1 = (x_{11} \quad x_{12} \quad \dots \quad x_{1p})^T$$
$$x_n = (x_{n1} \quad x_{n2} \quad \dots \quad x_{np})^T$$

and these observations fall into 2 classes, $y_1, y_2 \dots y_n \in \{-1, 1\}$ where -1 and 1 represent two classes respectively.

Let $x^* = (x_1^* \quad \dots \quad x_p^*)^T$ be a p-dimensional test vector

Goal is to classify x^* using feature measurement

Suppose it is possible to construct a hyperplane that seperates the training observations perfectly to their class labels. Then a seperating hyperplane has a property that

$$\beta^T x_i > 0 \quad \text{if} \quad y_i = 1$$
and $$\beta^T x_i < 0 \quad \text{if} \quad y_i = -1$$

Equivalently we can say that $y_i(\beta^T x_i) > 0 \quad \forall\ i = 1 \dots n$

Let $f(x^*) = \beta^T x^*$.

If $f(x^*) > 0$ classify x^* in class 1

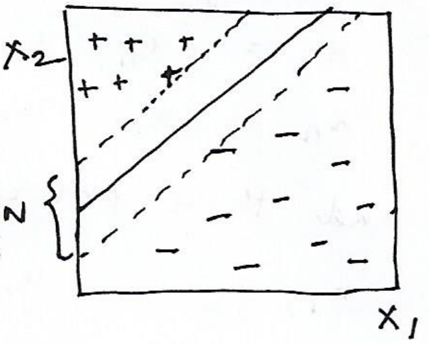
otherwise $f(x^*) < 0$ it ~~belog~~ belongs to class -1

More is the magnitude of $f(x^*)$, more certain we are certain about assignment

optimal seperating hyperplane OR Maximal margin classifier 2B

- The maximal margin hyperplane is the seperating hyperplane for which margin is largest ie hyperplane that has the farthest minimum distance to the training observations. This when used to classify test observations is called maxim margin classifier.

- $\boxed{\text{Maximal margin classifier}}$ depends only a small subset of observations (support vectors) and not entire training set.

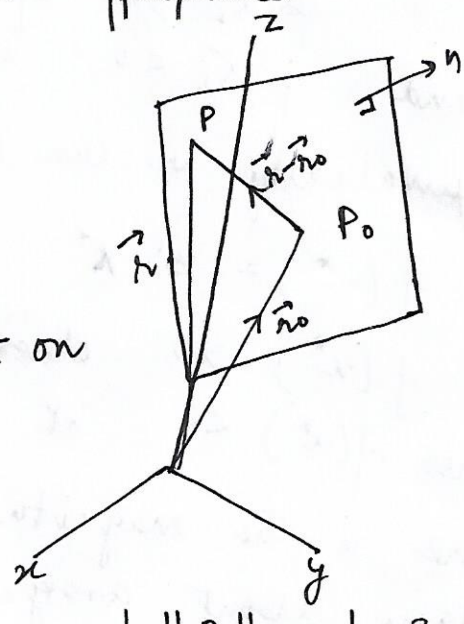


So $\boxed{\text{support vectors}}$ are vectors in p-dimensional space which "support" maximal margin classifier, that is, if one moves these, classifier moves

Hyperplane : strong line
Margin = distance betw two dotted lines

Note can result in overfitting
×

$\boxed{\text{BASIC-2}}$
Compute the distance from a point to hyperplane



1. consider hyperplane $\beta^T x = 0$ and point x_0.
2. Let x lie on the plane satisfying $\beta^T x = 0$
3. ~~consider~~ construct $x_0 - x$ which points x from x_0 and project it on β ($\because \beta$ is normal)
4. distance

$$d = \| \text{proj}_\beta (x_0 - x) \|$$

$$= \left\| \frac{(x_0 - x)\beta}{\beta \cdot \beta} \beta \right\| = \frac{|x_0\beta - x\beta| \, \|\beta\|}{\|\beta\|^2} = \frac{|x_0\beta - x\beta|}{\|\beta\|}$$

$$= \text{width of seperating plane lines}$$

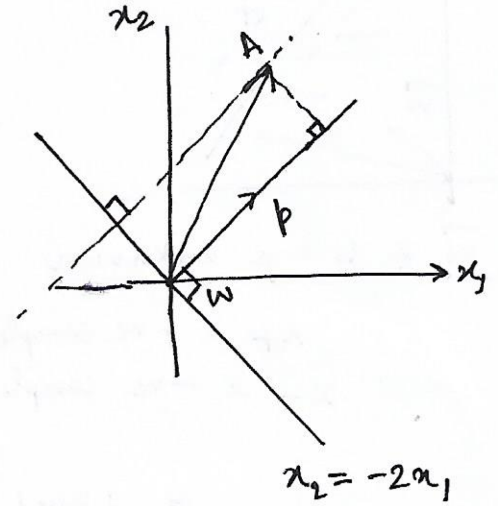
Eg Consider a hyperplane $x_2 = -2x_1$

Example

$\beta^T X = 0 \Leftrightarrow \beta^T = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

To compute distance b/w point A (3,4) and the hyperplane

This is the distance b/w A and its projection onto hyperplane



$\beta^T = (2, 1)$ $\quad \vec{a} = (3, 4)$ (Vector from origin to A)

$\|\vec{\beta}\| = \sqrt{2^2 + 1} = \sqrt{5}$

Let $\vec{u}$ be in direction of $\vec{\beta}$

$\vec{u} = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)$

$\vec{p}$ is the orthogonal projection of $\vec{a}$ onto $\vec{\beta}$

$\vec{p} = (\vec{u} \cdot \vec{a})\, \vec{u} = \left(\frac{6}{\sqrt{5}} + \frac{4}{\sqrt{5}}\right)\vec{u} = \frac{10}{\sqrt{5}}\vec{u}$

$= \left(\frac{20}{5}, \frac{10}{5}\right) = (4, 2)$

$\|\vec{p}\| = \sqrt{4^2 + 2^2} = \sqrt{20} = 2\sqrt{5}$

Margin $= 2\|\vec{p}\| = 4\sqrt{5}$

Let $\vec{x}_+$ and $\vec{x}_-$ be two points on the dotted lines

$\beta^T \vec{x}_+ = 1$ and $\beta^T x_- = -1$

$\Rightarrow \beta_1 x_{1+} + \dots + \beta_p x_{p+} = 1 - \beta_0$

and $\beta_1 x_{1-} + \dots + \beta_p x_{1-} = -1 - \beta_0$

$\Rightarrow \beta_1(x_{1+} - x_{1-}) + \dots + \beta_p(x_{1+} - x_{1-}) = 2$

$\Rightarrow \frac{\vec{W}}{\|\vec{W}\|}(X_+ - X_-)$

$\beta^T \vec{x}_+ = \beta_0 + \beta_1 \vec{x}_{1+} + \dots + \beta_p \vec{x}_p$

Let $W = (W_1 \dots W_p)$

$X_+ = x_{1+} \dots x_{p+}$

$X_- = x_{1-} \dots x_p$

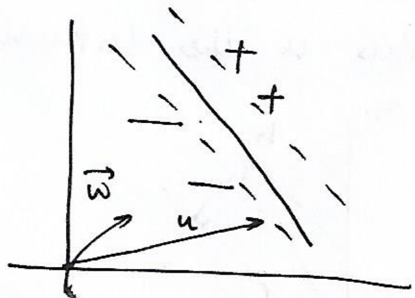
Consider decision rule

$\vec{w} \cdot \vec{u} + \vec{b} \geq 0$

- $\vec{w}$ ~~is~~ perpendicular to hyperplane
- $\vec{u}$ is a point ~~for~~ which we have to classify

A $\vec{w} \cdot \vec{u} \geq c$

WLOG $\vec{w}\,\vec{u} + b \geq 0$ Then + else − ~~else~~.



So we put constraints to find $\vec{w}$ & b

$\vec{w}$ & b are unknown

1 $\boxed{\vec{w} \cdot \vec{x}_+ + b \geq 1 \text{ (1)} \quad \text{and (2)} \quad \vec{w}\,\vec{x}_- + b \leq -1 \text{ (2)}}$

define y_i st $y_i = \begin{cases} +1 & \text{+ve samples} \\ -1 & \text{−ve samples} \end{cases}$

x_+ = +ve sam

x_- = −ve sam

(1) & (2) reduce to $y_i(\vec{w}\vec{x}_i + b) \geq 1$

B $\boxed{y_i(\vec{w}\vec{x}_i + b) - 1 = 0}$ for $\vec{x}_i$ for all samples on the dotted lines

$\boxed{\text{for step B}}$

Step B : Say $\vec{x}_+$ and $\vec{x}_-$ are points on dotted lines

width $\vec{x}_+ - \vec{x}_-$

$(\vec{x}_+ - \vec{x}_-) \dfrac{\vec{w}}{\|\vec{w}\|}$ = width of seperating lines

Using B, we have $\vec{w}\,\vec{x}_+ + b - 1 = 0$
& $-(\vec{w}\,\vec{x}_- + b) - 1 = 0$

$\Rightarrow \vec{w}\,\vec{x}_+ = 1 - b$ and $\vec{w}\,\vec{x}_- = -1 - b$

$\Rightarrow \vec{w}\,\vec{x}_+ - \vec{w}\,\vec{x}_- = \vec{w}(\vec{x}_+ - \vec{x}_-) = 1 - b + 1 + b = 2$

$\Rightarrow \vec{w}\,\vec{x}_+ - \vec{w}\,\vec{x}_- = 2$ Multiplying both sides by $\frac{1}{\|\vec{w}\|}$

$\Rightarrow \dfrac{1}{\|\vec{w}\|} \vec{w}(\vec{x}_+ - \vec{x}_-) = \dfrac{2}{\|\vec{w}\|}$

Thus, width of seperating hyperplane is $\dfrac{2}{\|\vec{w}\|}$

To get the widest seperating hyperplanes, maximize $\dfrac{2}{\|\vec{w}\|}$ or minimize $\|\vec{w}\|$

This is equivalent to ~~maximising~~ minimize $\frac{1}{2}\|W\|^2$.

which is a constraint quadratic programming problem with constraints & is solved using Lagrange's multiplier.

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i (\vec{w}_i \vec{x}_i + b) - 1 \right]$$

$\|\vec{w}\|^2 = \vec{w} \cdot \vec{w}$

$\frac{\partial \|\vec{w}\|^2}{\partial \vec{w}} = \vec{w} + \vec{w} = 2\vec{w}$

$$\frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i = 0$$

$$\Rightarrow \boxed{\vec{w} = \sum_{i=1}^{n} \alpha_i y_i \vec{x}_i}$$

ie $\vec{w}$ is a linear sum of samples.

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

$$L = \frac{1}{2}\|\vec{w}\|^2 - \vec{w}\cdot\vec{w} + \sum \alpha_i \Rightarrow \boxed{L = -\frac{1}{2}\|\vec{w}\|^2 + \sum_{i=1}^{n} \alpha_i}$$

$$\frac{\partial^2 L}{\partial \vec{w}^2} = 1 \qquad \frac{\partial^2 L}{\partial \vec{w}\,\partial b} = 0 \qquad \frac{\partial^2 L}{\partial b^2} = 0 \qquad \nabla^2 L = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix}$$

$$L = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \vec{x}_i \vec{x}_j + \sum_{i=1}^{n} \alpha_i \quad \text{— (4)}$$

Thus the decision rule is only dependent of $\vec{x}_i$ & $\vec{x}_j$

In case points are not linearly seperable, transform to higher dimension $\phi(\vec{x})$ and we need to maximize

$$\phi(\vec{x}_i)\,\phi(\vec{x}_j)$$

Kernels

Let $K(\vec{x}_i, \vec{x}_j)$ be the kernel func^n such that

$$K(\vec{x}_i\, \vec{x}_j) = \phi(\vec{x}_i)\phi(\vec{x}_j)$$

Thus using kernels one can produce non linear boundar[y] by constructing a linear bound[ary] in a large transformed version [of] the feature space