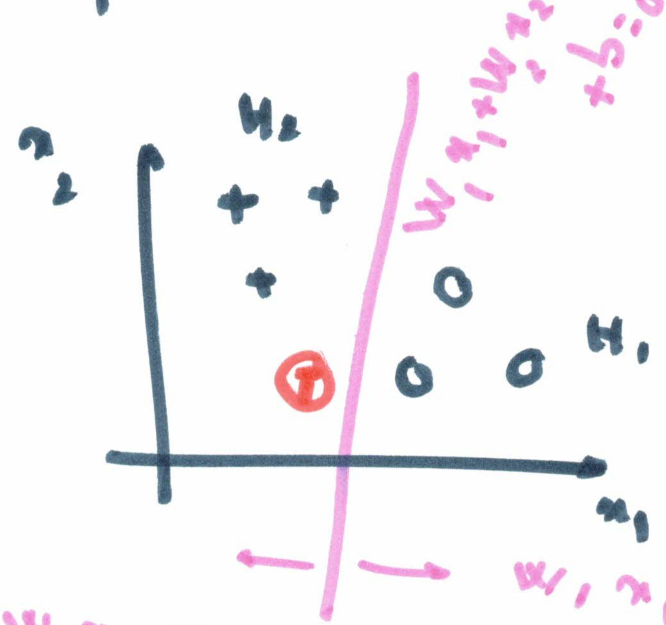
Support Vector Machine: (SVM):

Optimal Supervised classifier.



$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$; b

$\mathrm{sgn}(w_1 x_1^T + w_2 x_2^T + b)$

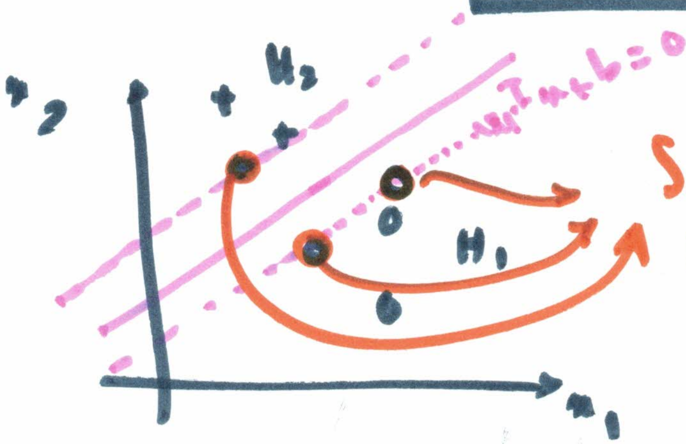
$w_1x_1+w_2x_2+b>0$ $\qquad$ $w_1x_1+w_2x_2+b<0$

equation of separating hyperplane:

$$\sum_i w_i x_i + b = 0 \quad \text{OR}$$

$$\boxed{w^T x + b = 0}$$

$d^i \in [+1, -1]$



Support vectors.

Data Points:

$x_s^1, x_s^2, \ldots, x_s^s$

$d^1, d^2, \ldots, d^s$

A problem:

$$w^T x + b = 0$$

$$(\alpha w^T)\, x + \alpha b = 0$$

So, w and b are Not unique because αw, αb can also be the solution.

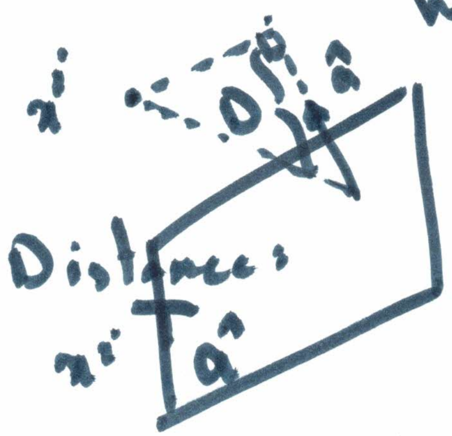
we need a normalization!

if x^i is a support vector (S.V.) then:

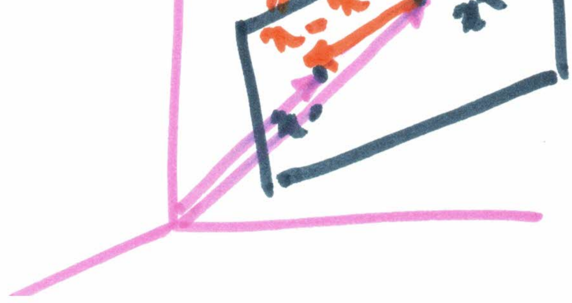
$$|w^T x^i + b| = 1$$

Distance:

vector w is orthogonal to the separating hyperplane.



Proof. pick two points on the separating hyperpla[ne] x', x''.



$$\left.\begin{aligned} w^T x' + b &= 0 \\ w^T x'' + b &= 0 \end{aligned}\right\} \Rightarrow$$

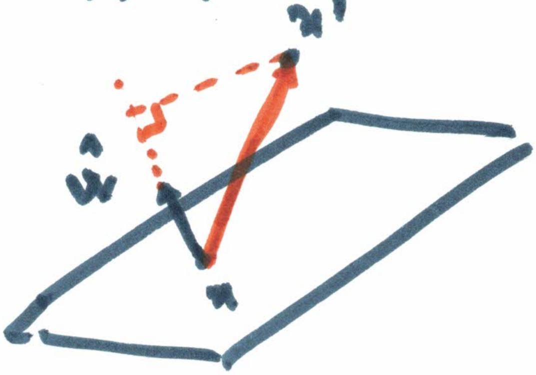
$$w^T(x' - x'') = 0 \Rightarrow w \perp x' - x''$$

This is true for any two points on the plane.
Therefore w is orthogonal to the plane.

7

$$\hat{w} = \frac{w}{\|w\|_2}$$

~~Distance between~~ if x is a point on the hyperplane and x^i is a data point.



Distance:

$$|(x^i - x)^T \hat{w}|$$

$$= \frac{|w^T x^i - w^T x|}{\|w\|_2}$$

$$= \frac{|w^T x^i + b - w^T x - b|}{\|w\|_2}$$

x is on the plane.

Assume x^i is a S.V.

$$= \frac{|\overbrace{(w^T x^i + b)}^{+1,\,-1} - \overbrace{(w^T x + b)}^{=0}|}{\|w\|_2}$$

Distance for support vectors is:

$$\frac{1}{\|w\|_2}$$

SVM Problem:

8

$$\text{Max.} \quad \frac{1}{\|W\|_2}$$

$$\text{s.t.} \quad \underset{\forall x^i}{\text{Min.}} \; |w^T x^i + b| = 1$$

we need to change:

Primal SVM optimization.

$$\text{Min.} \quad \frac{1}{2} w^T w$$

$$\text{s.t.} \quad d^i (w^T x^i + b) \geq 1$$

$$i = 1, \ldots, \xi$$

$$\mathcal{L}(w, b; \alpha_1, \ldots, \alpha_\xi) =$$

$$\frac{1}{2} w^T w - \sum_{i=1}^{\xi} \alpha_i \left\{ d^i (w^T x^i + b) - 1 \right\}$$

we know: $\alpha_i \geq 0$

* we expect many of these α_i's to be zero!

$$\nabla_w \mathcal{L} = 0 \Rightarrow w - \sum_{i=1}^{\xi} \alpha_i d^i x^i = 0$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow - \sum_{i=1}^{\xi} \alpha_i d^i = 0$$

$$w = \sum_{i=1}^{\xi} \alpha_i d^i x^i$$

$$\sum_{i=1}^{\xi} \alpha_i d^i = 0$$

$$\mathcal{L} = \frac{1}{2} w^T w - \sum_{i=1}^{\xi} \alpha_i \left\{ d^i \left(w^T x^i + b \right) - 1 \right\}$$

$$\mathcal{L} = \frac{1}{2} \left(\sum_{i=1}^{\xi} \alpha_i d^i x^i \right)^T \left(\sum_{j=1}^{\xi} \alpha_j d^j x^j \right)$$

$$- \sum_{i=1}^{\xi} \alpha_i \left\{ d^i \left\{ \left(\sum_{j=1}^{\xi} \alpha_j d^j x^j \right)^T x^i + b \right\} - 1 \right\}$$

$$\mathcal{L}_d = \sum_{i=1}^{\xi} \alpha_i - \frac{1}{2} \sum_{i=1}^{\xi} \sum_{j=1}^{\xi} \alpha_i \alpha_j d^i d^j x^{i^T} x^j$$

This is $\inf_{w} \mathcal{L} = g(\alpha)$

$$g(\alpha) = \sum_{i=1}^{\xi} \alpha_i - \frac{1}{2} \sum_{i=1}^{\xi} \sum_{j=1}^{\xi} \alpha_i \alpha_j d^i d^j x^{i^T} x^j$$

dual Problem: $\max_{\alpha} \; g(\alpha)$

$\alpha_i \geq 0$

So:

$$\text{Max.} \; \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \alpha_i \alpha_j d^i d^j x^{i^T} x^j$$

$$\text{s.t.} \quad \alpha_i \geq 0 \; , \quad i = 1, \dots, \ell \qquad \sum_{i=1}^{\ell} \alpha_i d^i = 0$$

<u>reformulate:</u>

$$\min_{\alpha_i} \; \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \alpha_i \alpha_j d^i d^j x^{i^T} x^j - \sum_{i=1}^{\ell} \alpha_i$$

$$\text{s.t.} \quad \alpha_i \geq 0 \; , \quad i = 1, \dots, \ell \; , \quad \sum_{i=1}^{\ell} \alpha_i d^i = 0$$

<u>Can be written as:</u>

$$\min_{\alpha_i} \; \frac{1}{2} \alpha^T \begin{bmatrix} d^1 d^1 x^{1^T} x^1 & d^1 d^2 x^{1^T} x^2 & \cdots & d^1 d^{\ell} x^{1^T} x^{\ell} \\ \vdots & & \ddots & \vdots \\ d^{\ell} d^1 x^{\ell^T} x^1 & \cdots & & d^{\ell} d^{\ell} x^{\ell^T} x^{\ell} \end{bmatrix} \alpha - I^T \alpha$$

$$\text{s.t.} \quad d^T \alpha = 0 \quad ; \quad \alpha_i \geq 0 \qquad I = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\min_{\alpha} \ \frac{1}{2}\alpha^T Q \alpha - \bar{I}^{\,T} \alpha$$

$$\text{s.t.} \quad d^T \alpha = 0$$

$$\alpha_i \geq 0$$

dual Problem

4

Matlab: quadprog commad.

$$Q = \begin{bmatrix} d^1 d^1 x^{1T} x^1 & d^1 d^2 x^{1T} x^2 & \cdots & d^1 d^f x^{1T} x^f \\ d^2 d^1 x^{2T} x^1 & d^2 d^2 x^{2T} x^2 & \cdots & d^2 d^f x^{2T} x^f \\ \vdots & & & \vdots \\ \cdot & \cdot & \cdots & d^f d^f x^{fT} x^f \end{bmatrix}_{f \times f}$$

we can write Q as:

$$Q = \begin{bmatrix} d_1 x_1^1 & d_1 x_2^1 & \cdots & d_1 x_n^1 \\ d_2 x_1^2 & d_2 x_2^2 & \cdots & d_2 x_n^2 \\ \vdots & & & \vdots \\ \cdot & \cdot & \cdots & d_f x_n^f \end{bmatrix} \begin{bmatrix} d_1 x_1^1 & \cdots & d_f x_1^f \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \cdot & \cdots & d_f x_n^f \end{bmatrix}$$

$$= B^T B$$

so Q is P.D.

so dual Problem is convex.
This is a Quadratic Problem.

Solve this problem to find optimal α.

$$\boxed{\begin{aligned} W &= \sum_{i=1}^{\xi} \alpha_i d^i x^i \\ &= \sum_{s.v.} \alpha_i d^i x^i \end{aligned}}$$

How can I find b?

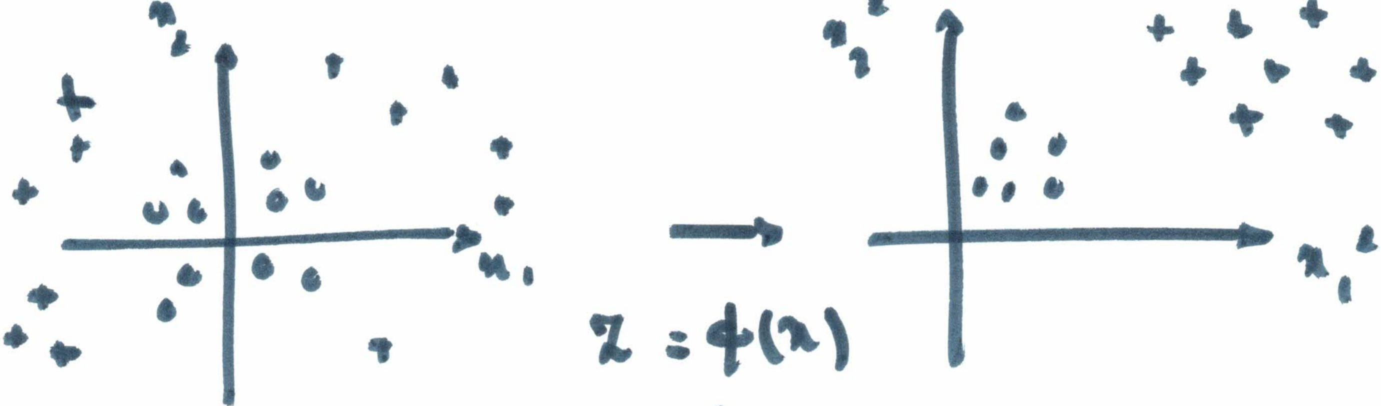
Use normalization. Pick one of your s.v. s:

$$d^i \left(W^T x^i + b \right) = 1$$

if x^i is a s.v.

Solve this equation for b

$$W^T x^i + b = d^i \Rightarrow \boxed{b = d^i - W^T x^i}$$



Map data to higher dimensions.

$$(x_1, x_2) \longrightarrow (x_1, x_2, x_1^2, x_2^2)$$

by going to z domain where $z = \phi(x)$ dimension of the SVM optimization does NOT change.

Only the new Q Matrix is:

$$Q = \begin{bmatrix} d^1 d^1 z^{1T} z^1 & d^1 d^2 z^{1T} z^2 & \cdots & d^1 d^{\xi} z^{1T} z^{\xi} \\ d^2 d^1 z^{2T} z^1 & d^2 d^2 z^{2T} z^2 & \cdots & \vdots \\ \vdots & & & \\ \cdot & \cdots & \cdot & d^{\xi} d^{\xi} z^{\xi T} z^{\xi} \end{bmatrix}_{\xi \times \xi}$$

$$w = \sum_{i=1}^{\xi} \alpha_i d^i z^i = \sum_{S.V.} \alpha_i d^i z^i$$

$$b = d^i - w^T z^i = d^i - \sum_j \alpha_j d^j z^{jT} z^i$$

$$\mathrm{sgn}\left(w^T x + b\right)$$

$$\mathrm{sgn}\left(\left(\sum \alpha_i d^i z^i\right)^T z + b\right)$$

$$\mathrm{sgn}\left(\sum \alpha_i d^i z^{i^T} z + b\right)$$

(14)

it seems that what we alway need is not the z values but all $z^{i^T} z^j$ s.

Kernels

$$K(x, x') = z^T z'$$

Ex: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\quad z = \phi(x) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix}$

$x' = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}$

$K(x, x') = z^T z'$ $\quad z' = \phi(x') = \begin{bmatrix} 1 \\ x_1' \\ x_2' \\ \vdots \\ x_1' x_2' \end{bmatrix}$

$$= 1 + x_1 x_1' + x_2 x_2'$$
$$+ x_1^2 x_1'^2 + x_2^2 x_2'^2 + x_1 x_1' x_2 x_2'$$

Now think about this function:

$$K(x, x') = (1 + x^T x')^2$$

$$= \left(1 + [x_1 \; x_2] \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}\right)^2$$

$$= (1 + x_1 x_1' + x_2 x_2')^2$$

$$= 1 + x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x_1' + 2x_2 x_2' + 2x_1 x_1' x_2 x_2'$$

This can be the inner product if

$$Z_{New} = \phi_{New}(x) = \begin{bmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ x_1^2 \\ x_2^2 \\ \sqrt{2}x_1 x_2 \end{bmatrix}$$

Then: $K(x, x') = (1 + x^T x')^2 = Z_{New}^T Z_{New}'$

Therefore $K(x, x') = (1 + x^T x')^2$ is a Kernel.

Polynomial Kernel:

15

$$K(x, x') = (1 + x^T x')^Q$$

$Q = 2, 3, \ldots$

in the SVM Problem:

Matrix Q:

$$Q = \begin{bmatrix} d^1 d^1 K(x^1, x^1) & d^1 d^2 K(x^1, x^2) \ldots d^1 d^N K(x^1, x^N) \\ & \\ & \end{bmatrix}$$

Radial Basis Kernel: (RBK) Function (RBF)

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right)$$

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Regular SVM:

$$\operatorname{sgn}(w^T x + b) \quad ; \quad w = \sum_{S.V.} \alpha_i d^i x^i$$

$$b = d^j - w^T x^j$$

x^j being a S.V.

(17)

in Kernel SVM:

$$\operatorname{sgn}(w^T z + b)$$

$$w = \sum_{i=1}^{s} \alpha_i d^i z^i$$

$$\operatorname{sgn}\left(\sum_{i=1}^{s} \alpha_i d^i z^{i^T} z + b \right)$$

$$\operatorname{sgn}\left(\sum_{i=1}^{s} \alpha_i d^i K(x_i, x) + b \right)$$

$$b = d^j - w^T z^j = d^j - \sum_{i=1}^{s} \alpha_i d^i z^{i^T} z^j$$

$$= d^j - \sum_{i=1}^{s} \alpha_i d^i K(x_i, x_j)$$

CVX toolbox in Matlab