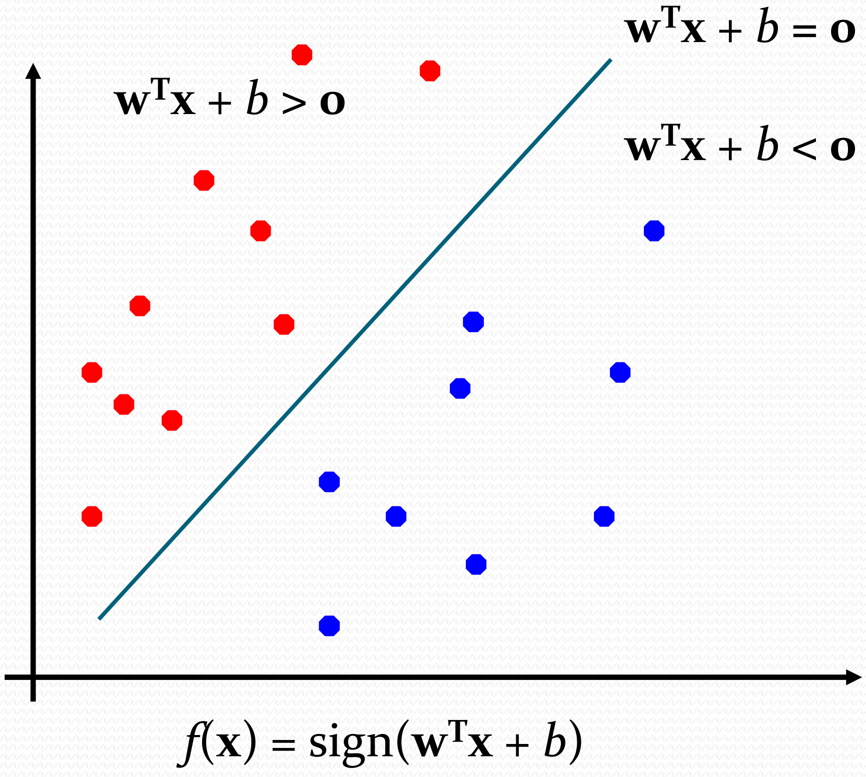
Support Vector Machine

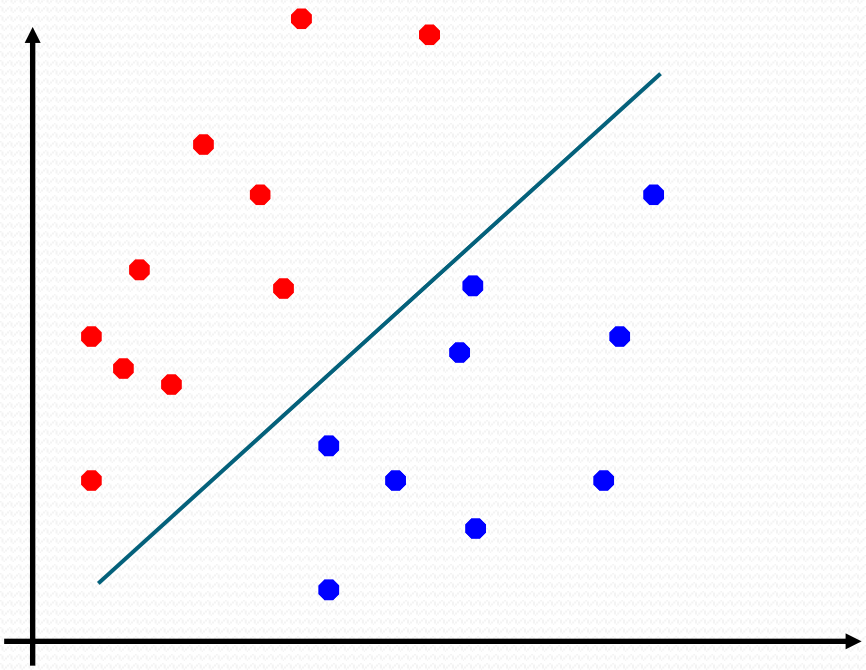
1. Tổng quan

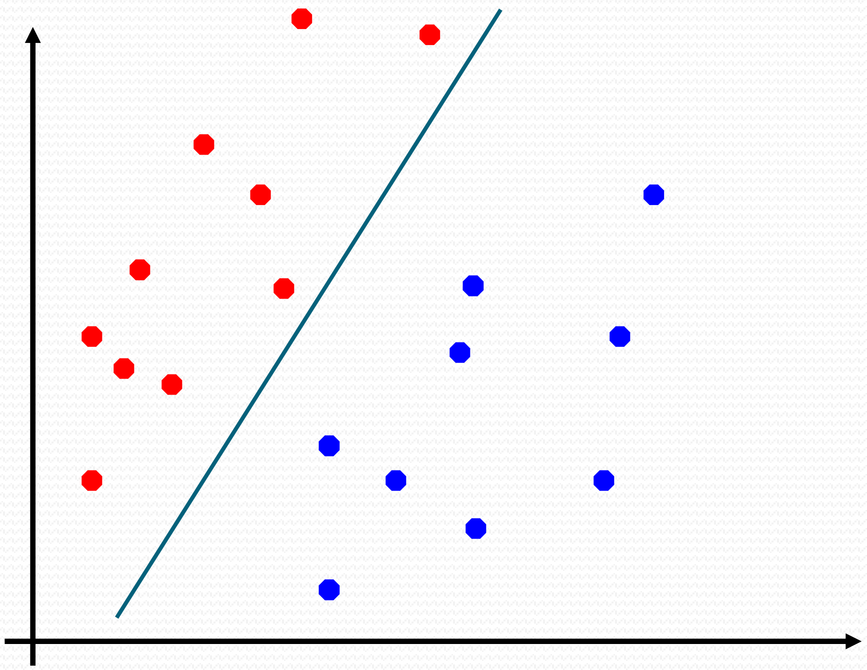


- Bài toán SVC: Đối với bộ dữ liệu được phân chia tuyến tính, tìm hyperplane tốt nhất để phân chia bộ dữ liệu thành 2 lớp riêng biệt.

1. Tổng quan

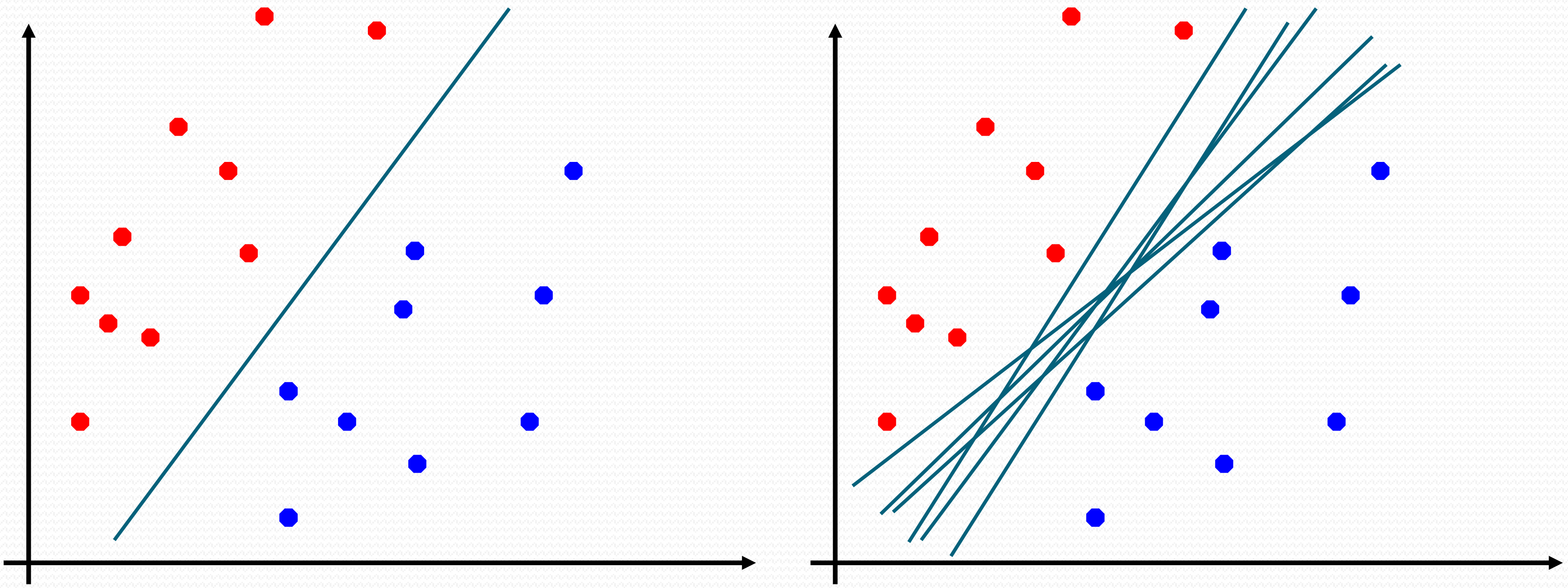
- Siêu phẳng nào là tốt nhất?





1. Tổng quan

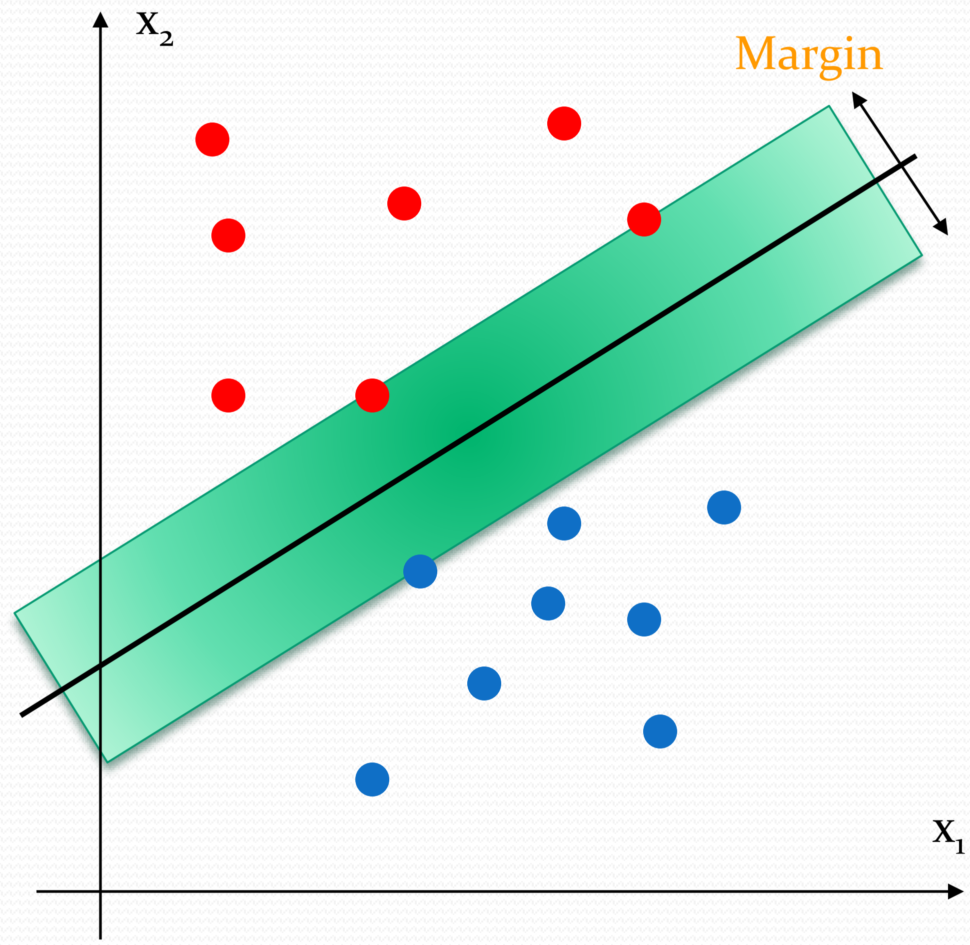
- Siêu phẳng nào là tốt nhất?



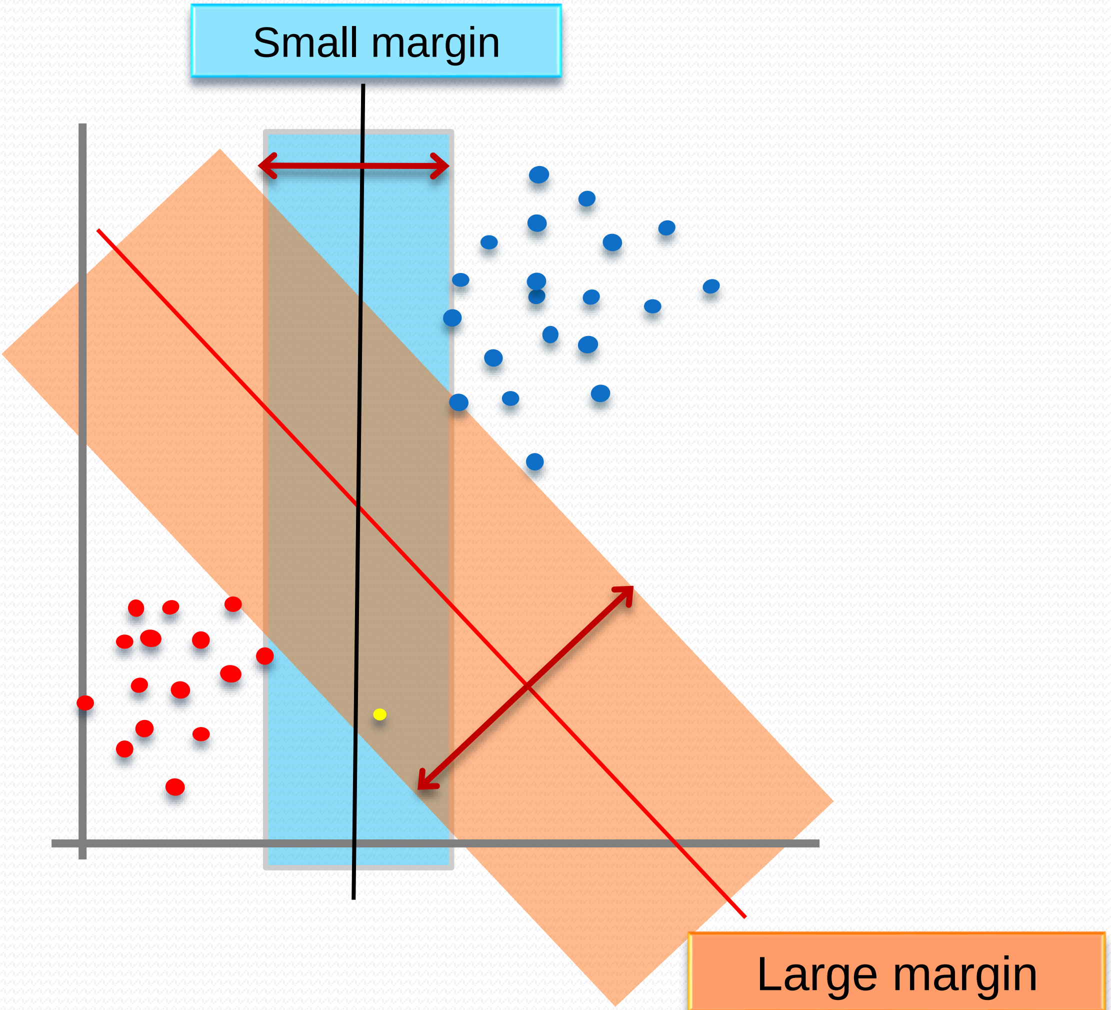
1. Tổng quan

- Biên (margin): là khoảng cách ngắn nhất giữa điểm dữ liệu gần hyperplane nhất đến hyperplane.
- Phương trình siêu phẳng

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



1. Tổng quan



2. Support Vector Classifer

- Cho tập điểm dữ liệu

$$\{(\mathbf{x}_i, y_i)\},\ i = 1, 2, \cdots, n$$

Trong đó:

Với: $y_i = +1, \quad \mathbf{w}^T\mathbf{x}_i + b > 0$

Với: $y_i = -1, \quad \mathbf{w}^T\mathbf{x}_i + b < 0$

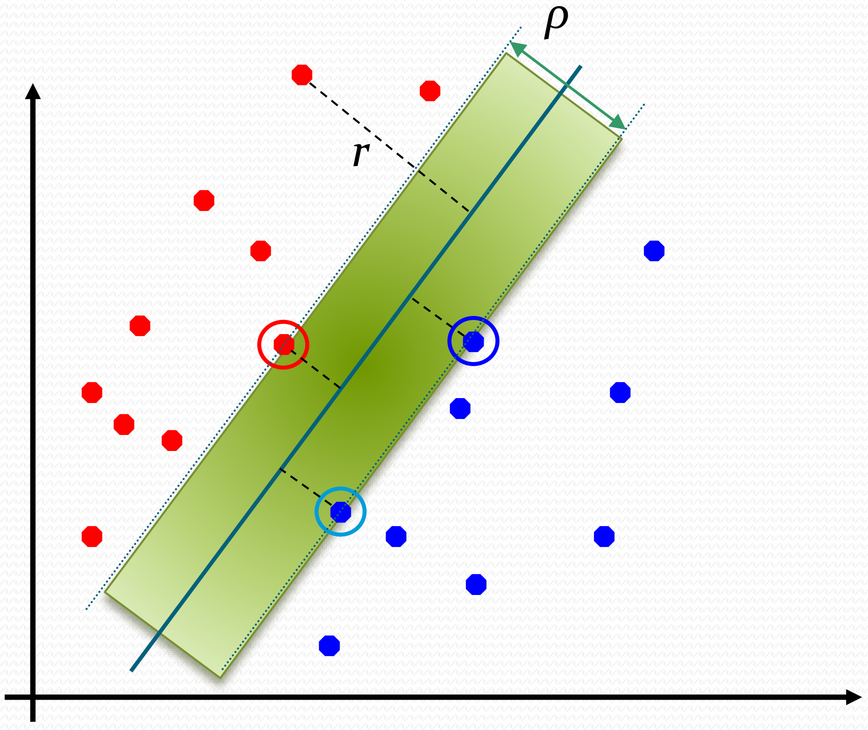
- Tương đương với

Với: $y_i = +1, \quad \mathbf{w}^T\mathbf{x}_i + b \geq 1$

Với: $y_i = -1, \quad \mathbf{w}^T\mathbf{x}_i + b \leq -1$

- Kết hợp 2 bpt trên:

$$y_i(wx_i + b) \geq 1$$

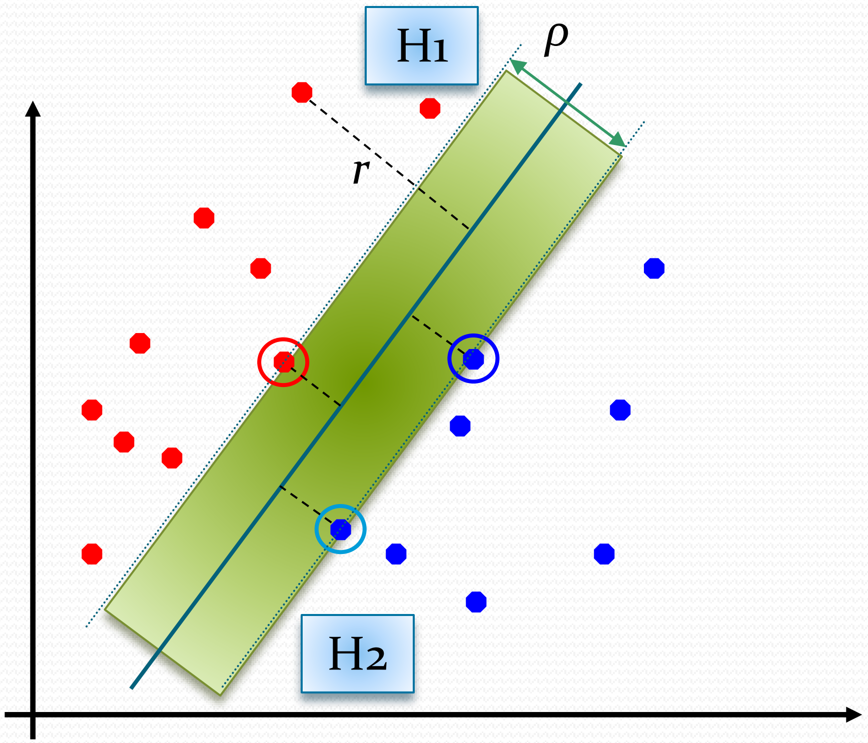


2. Support Vector Classifer

- Support vector

  x_i: $\boldsymbol{w} \,.\, \boldsymbol{x_i} + b = +1 \Rightarrow H1$

  x_i: $\boldsymbol{w} \,.\, \boldsymbol{x_i} + b = -1 \Rightarrow H2$



2. Support Vector Classifer

- Độ rộng lề

  d1(O,H1) = (1-b)/||w||

  d2(O,H2) = (-1-b)/||w||

  => d=|d1-d2| = **2/||w||**

- Để tìm siêu phẳng lề cực đại

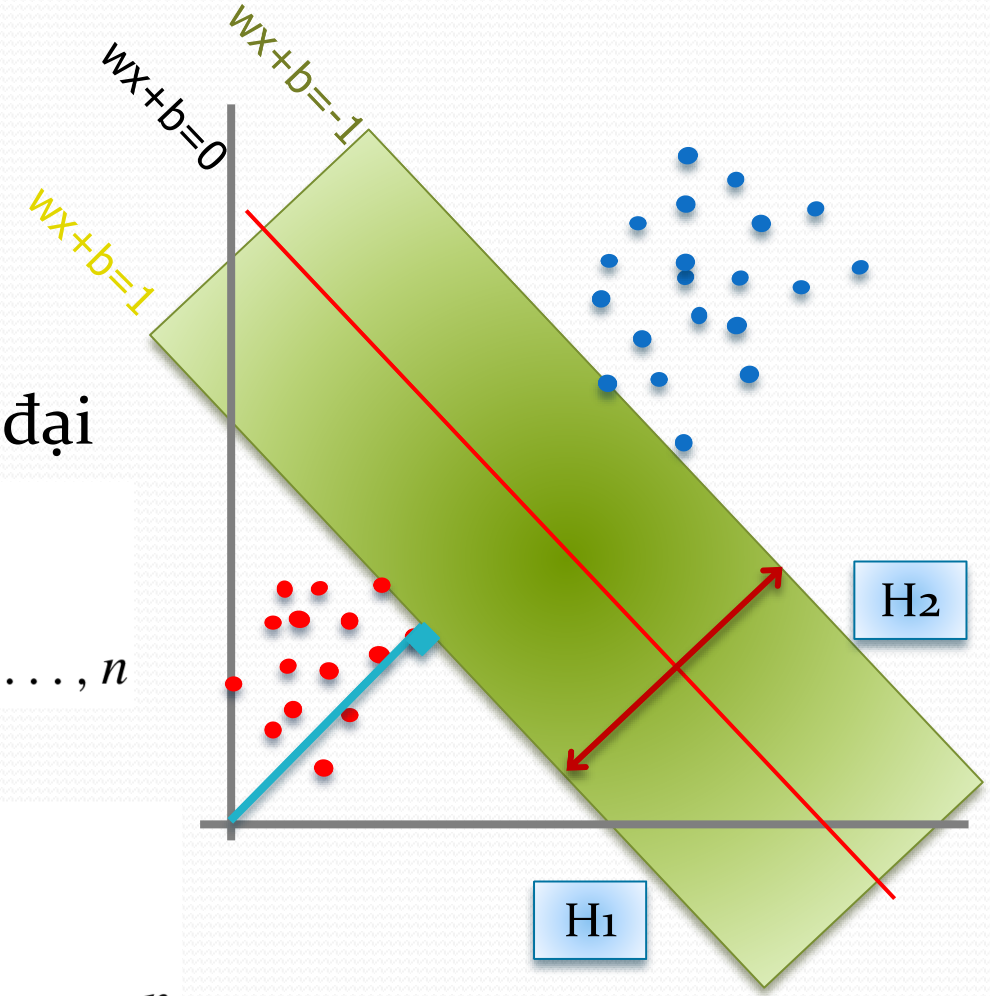
$$\max_{\mathbf{w},b} \frac{2}{\|\mathbf{w}\|}$$

$$s.t.\ y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1, \quad i = 1, \ldots, n$$

⇔

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$

$$s.t.\ y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \quad i = 1, \ldots, n$$



2. Support Vector Classifer

- Bài toán tìm cực tiểu

$$\min_{\mathbf{w},b} \frac{1}{2} \|\mathbf{w}\|^2$$

$$s.t.\ y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, \ldots, n$$

- Hàm Lagrange

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

αi: hệ số nhân Lagrange

2. Support Vector Classifer

- Bài toán tìm cực tiểu

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

- Cực trị của L xảy ra tại w và b sao cho

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

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$

2. Support Vector Classifer

- Bài toán tìm cực tiểu

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

- Viết lại

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$s.t. \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \quad i = 1, \ldots, n$$

2. Support Vector Classifer

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$s.t. \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \quad i = 1, \ldots, n$$

- Giải bài toán này ta tìm được α_i, từ đó tìm được w từ công thức:

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$$

2. Support Vector Classifer

- Để tính b: sử dụng điều kiện KKT cho bài toán gốc

$$\alpha_i \left[y_i \left(\mathbf{w}^T \mathbf{x}_i + b \right) - 1 \right] = 0, \quad i = 1, \ldots, n$$

- α_i = 0: x_i không nằm trên siêu phẳng, biên H1 hay H2

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$$

- α_i > 0: x_i nằm trên H1 hay H2, xi được gọi là support vector.

2. Support Vector Classifer

- Để tính b chọn $\alpha_i > 0$:

$$\alpha_i \left[y_i \left(\mathbf{w}^T \mathbf{x}_i + b \right) - 1 \right] = 0, \quad i = 1, \ldots, n$$

Nên: $\left[y_i \left(\mathbf{w}^T \mathbf{x}_i + b \right) - 1 \right] = 0$

$$b^* = 1 - \mathbf{w}^{*T} \mathbf{x}_s$$

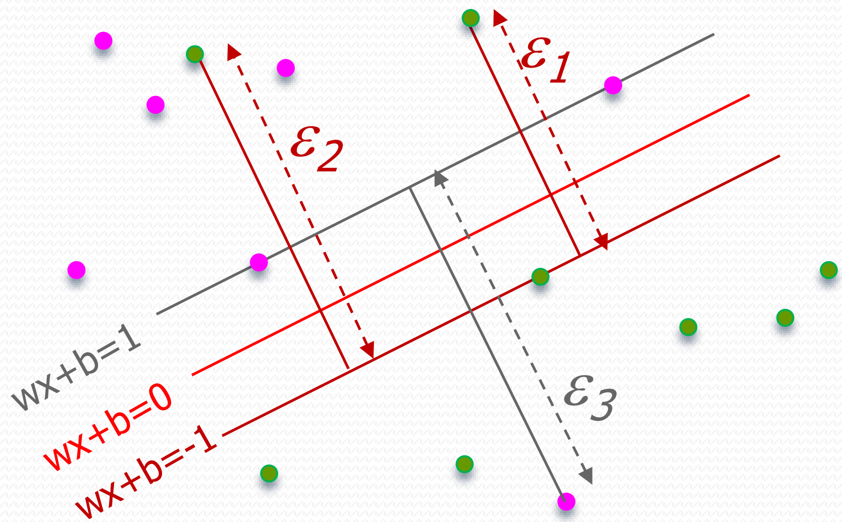
với: $y_i = 1$

- Nhận dạng: cần nhận dạng x* thuộc (1 hay -1)

$$y^* = f(x^*) = sign(w^T x^* + b)$$

$$= sign(\sum_{i=1}^{n} \alpha_i y_i x_i x^* + b)$$

2. SVC + Soft margin



Vấn đề: Khi sự phân bố của dữ liệu không phân chia một cách tuyến tính, dẫn đến không thể tìm một hyperplane.

Soft margin:

- Chấp nhận 1 số mẫu phân lớp sai

-N: số điểm dữ liệu (x_k,y_k) với y_k=+/-1
$-\varepsilon_i$: khoảng cách từ điểm i tới lề đúng của nó
-C: tham số hiệu chỉnh độ lỗi

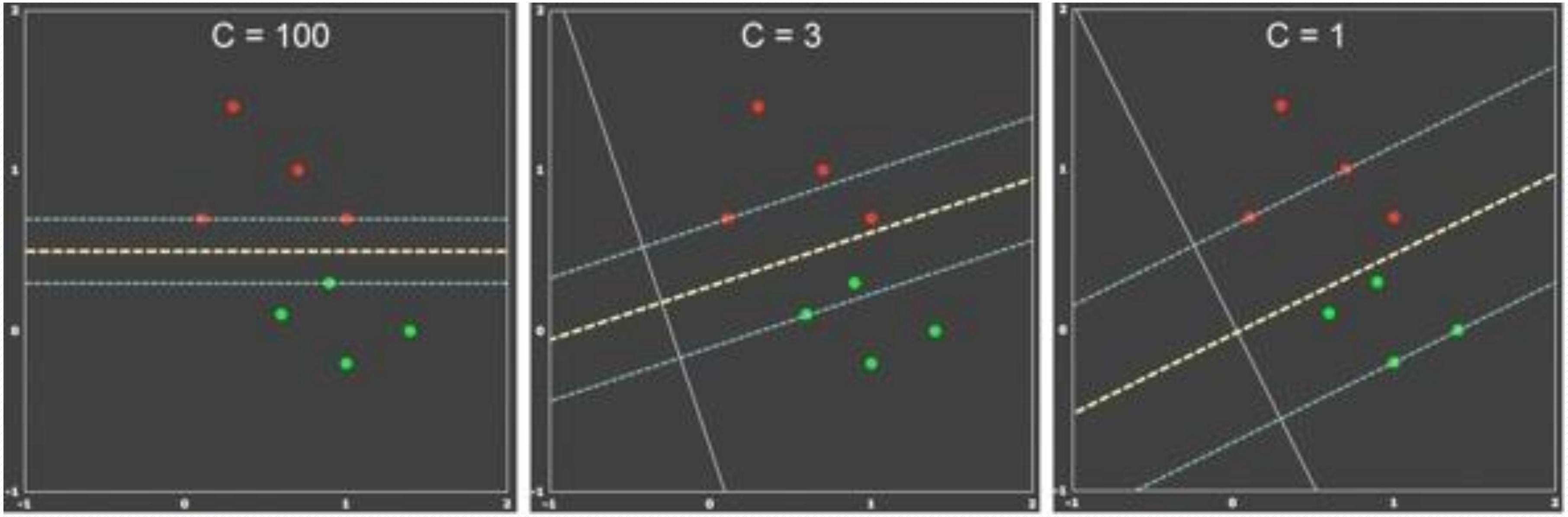
2. SVC + Soft margin

- Để giải quyết bài toán này, SVC + Soft margin chấp nhận độ lỗi nhất định khi tìm hyperplane (giảm ràng buộc ở bài toán gốc – primal problem):

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$s.t. \quad y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, n$$

2. SVC + Soft margin



2. SVC + Soft margin

Tương nhự như SVC, áp dụng phương pháp nhân tử Lagrange ta có được bài toán đối ngẫu:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$s.t. \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, \quad i = 1, \ldots, n$$

2. SVC + Soft margin

Điều kiện bổ sung Karush-Kuhn-Tucker

$$\alpha_i \left[y_i \left(\mathbf{w}^T \mathbf{x}_i + b \right) - 1 + \xi_i \right] = 0, \quad i = 1, \ldots, n$$

$$\gamma_i \xi_i = 0, \quad i = 1, \ldots, n$$

Do tại điểm cực trị của hàm Lagrange chúng ta muốn $\xi_i = 0$, nên ta có điều kiện

$$\alpha_i + \gamma_i = C$$

Hay

$$\xi_i = 0 \quad if \quad \alpha_i < C$$

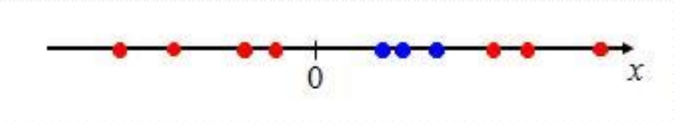
2. SVC + Soft margin

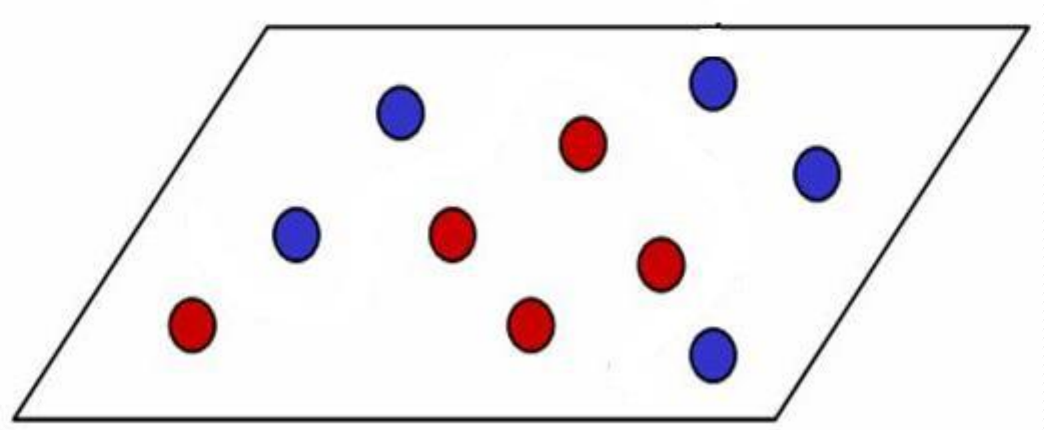
Từ điều kiện trên ta tìm được vector $\mathbf{w}^*$ tối ưu

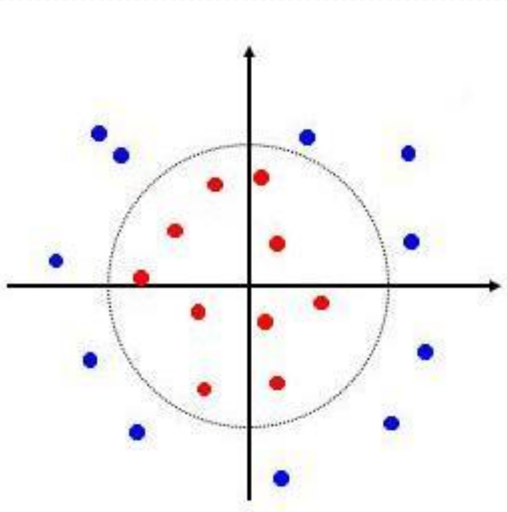
$$\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i^* y_i \mathbf{x}_i$$

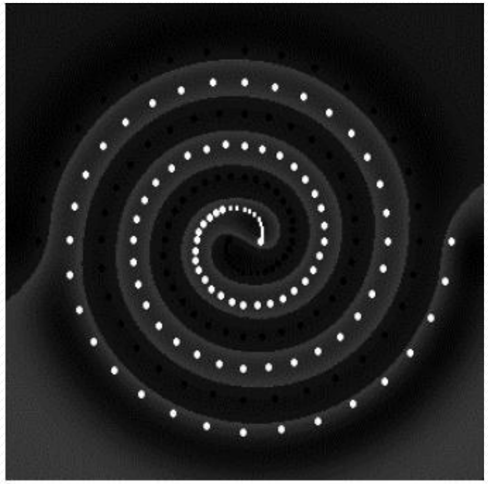
Khi tính được $\mathbf{w}^*$, chúng ta tính giá trị b^* tương tự như ở bài toán SVC bằng bất kỳ dữ liệu nào có $\xi_i = 0$

3.1 Kernel là gì ?



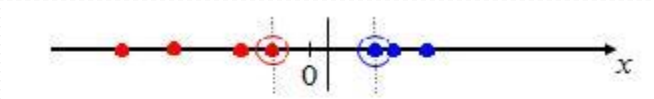




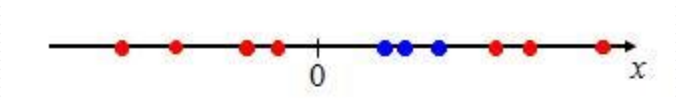


3.1 Kernel là gì ?

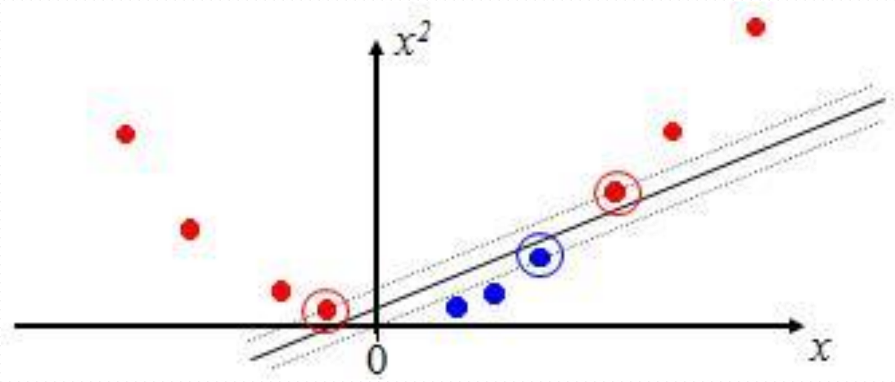
- * Với tập dữ liệu phân chia tuyến tính được :



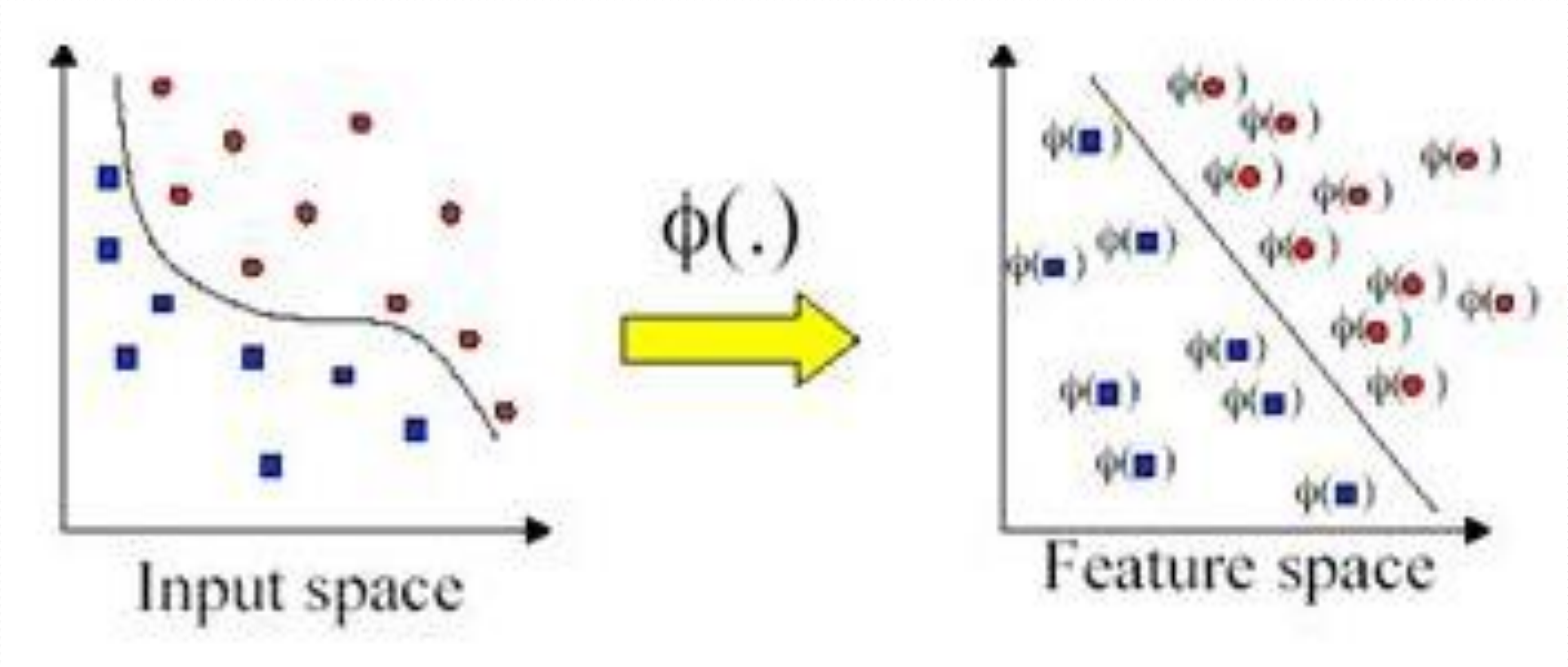
- * Với tập dữ liệu phi tuyến :



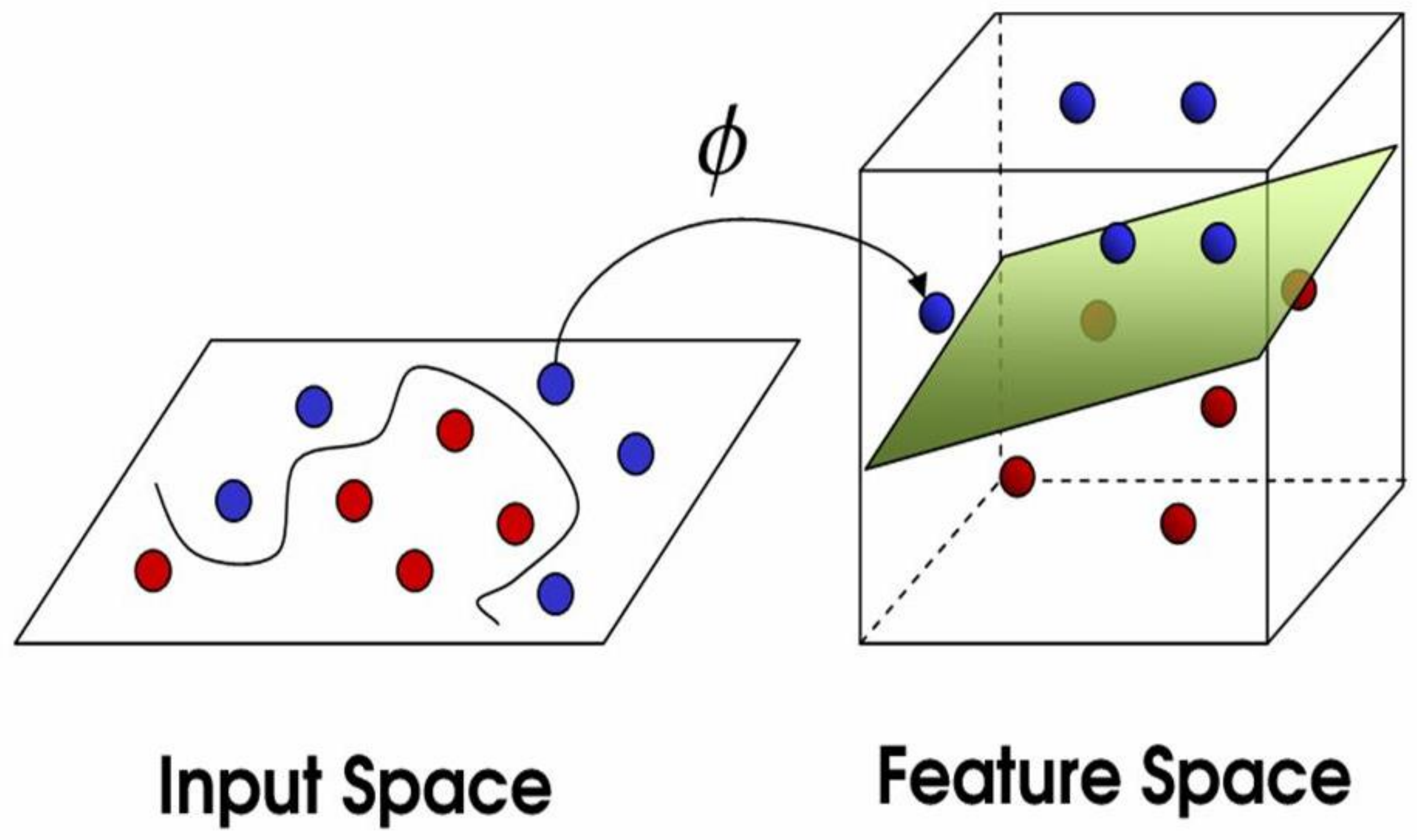
- * Cách giải quyết :



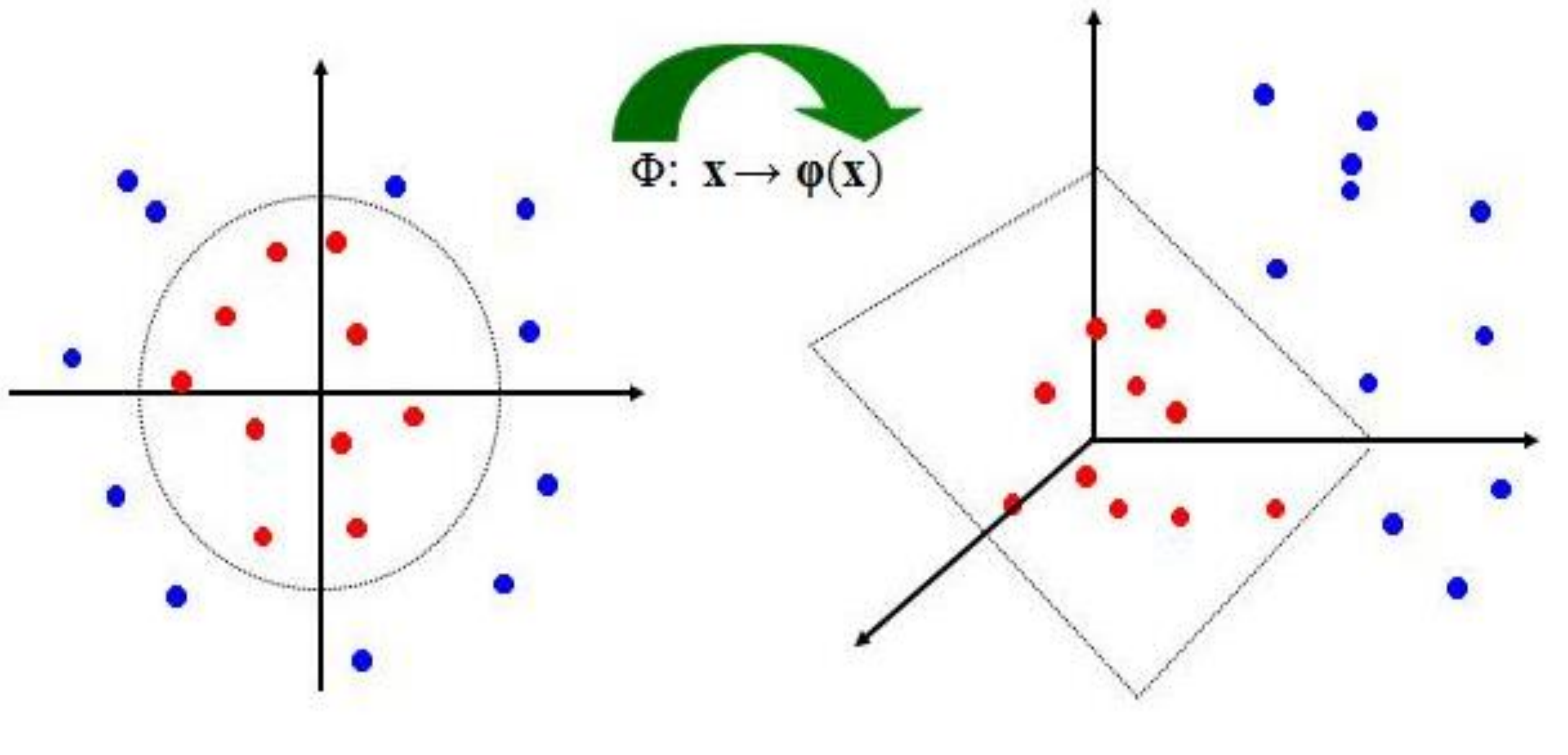
3.1 Kernel là gì ?



3.1 Kernel là gì ?



3.1 Kernel là gì ?



3.1 Kernel là gì ?

- Định nghĩa hàm Kernel :
  - <u>Định nghĩa :</u> Cho X, Y là 2 tập cho trước, hàm f được gọi là kernel ánh xạ từ X vào Y khi thỏa điều kiện sau :
    - - x1, x2 thuộc X bằng nhau khi và chỉ khi f(x1)=f(x2) (với f(x1), f(x2) là 2 phần tử thuộc Y)

  Công thức f : X ->Y

$$\ker(f) = \{(x_1, x_2) \in X \times X : f(x_1) = f(x_2)\}$$

3.1 Kernel là gì ?

Định lý: Một hàm $k : x \times x \rightarrow y$ có thể được viết là $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$ khi $\Phi(x)$ là một ánh xạ đặc điểm $x \rightarrow \Phi(x) \in F$ nếu k(x,y) thỏa mãn thuộc tính có thể xác định.

+ Điều kiện Mercer

$K(x, y) = \sum_i \Phi(x)_i \Phi(y)_i$ nếu và chỉ nếu với mọi g(x) thỏa mãn

$\int g(x)^2 dx$ là hữu hạn thì $\int K(x, y) g(x) g(y) dx dy \geq 0$

3.2 Một số loại Kernel

- Linear: $K(\mathbf{x}_i,\mathbf{x}_j)=\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j$
- Polynomial of power p: $K(\mathbf{x}_i,\mathbf{x}_j)= \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j{}^p$
  $K(\mathbf{x}_i,\mathbf{x}_j)=(1+\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)^p$
- Gaussian (radial-basis function): $K(\mathbf{x}_i,\mathbf{x}_j)= e^{-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\sigma^2}}$

  σ là thông số điều khiển độ rộng của hàm kernel xung quanh x_i.

3.2 Một số loại Kernel

- Ví dụ :

  Xét vector $\mathbf{x}=[x_1\ \ x_2]$, $z=[z_1\ \ z_2]$; tính $K(\mathbf{x},\mathbf{z})=(\mathbf{x}^T\mathbf{z})^2$,

  Cần tìm $\boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(z)$ ứng với $K(\mathbf{x},z)$:

$$\langle x, z\rangle^2 = (x_1z_1 + x_2z_2)^2 =$$

$$= x_1^2z_1^2 + x_2^2z_2^2 + 2x_1z_1x_2z_2 =$$

$$= \left\langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (z_1^2, z_2^2, \sqrt{2}z_1z_2) \right\rangle =$$

$$= \langle \phi(x), \phi(z) \rangle$$

3.3 Xây dựng Kernel

- Xây dựng hàm $\phi()$để kiểm tra $K(\mathbf{x}_i,\mathbf{x}_j)= \phi(\mathbf{x}_i)^{\mathbf{T}}\phi(\mathbf{x}_j)$ là không cần thiết.

- Hàm xác định nửa ma trận đối xứng positive được xác định bởi ma trận Gram :

3.3 Xây dựng Kernel

K=

$K(\mathbf{x}_1,\mathbf{x}_1)$	$K(\mathbf{x}_1,\mathbf{x}_2)$	$K(\mathbf{x}_1,\mathbf{x}_3)$	…	$K(\mathbf{x}_1,\mathbf{x}_n)$
$K(\mathbf{x}_2,\mathbf{x}_1)$	$K(\mathbf{x}_2,\mathbf{x}_2)$	$K(\mathbf{x}_2,\mathbf{x}_3)$		$K(\mathbf{x}_2,\mathbf{x}_n)$
…	…	…	…	…
$K(\mathbf{x}_n,\mathbf{x}_1)$	$K(\mathbf{x}_n,\mathbf{x}_2)$	$K(\mathbf{x}_n,\mathbf{x}_3)$	…	$K(\mathbf{x}_n,\mathbf{x}_n)$

3.3 Xây dựng Kernel

- Xét trên tập Kernel đóng. Nếu K & K’ là kernel, ta có :
  - K + K’ cũng là kernel
  - cK là kernel ,(c>o)
  - aK + bK’ là kernel, (a,b>o)
  - V.v.....

  - Có thể xây dựng kernel phức tạp từ những kernel đơn giản.

4. Support Vector Regression

Áp dụng Support Vector Machine vài bài toán hồi quy: Support Vector Regression (SVR)

- Hồi quy tuyến tính (Linear Regression)

4. SVR-Ý Tưởng

- Xét 1 tập dữ liệu:

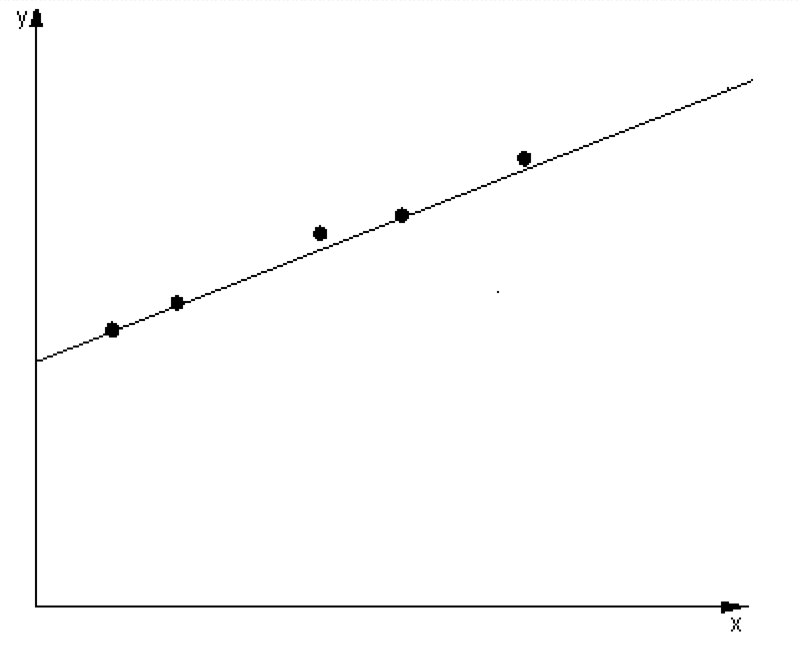
$$D = \{(x_1, y_1), (x_2, y_2) \ldots (x_l, y_l)\}, \quad x \in R^n, y \in R$$

- Tìm hàm f(x) sao cho xấp xỉ tập trên

- Ví dụ : hàm f tuyến tính có dạng

$$f(x) = < w, x > + b, \quad \mathrm{w} \in R^n, b \in R$$

5. SVR- Ý Tưởng

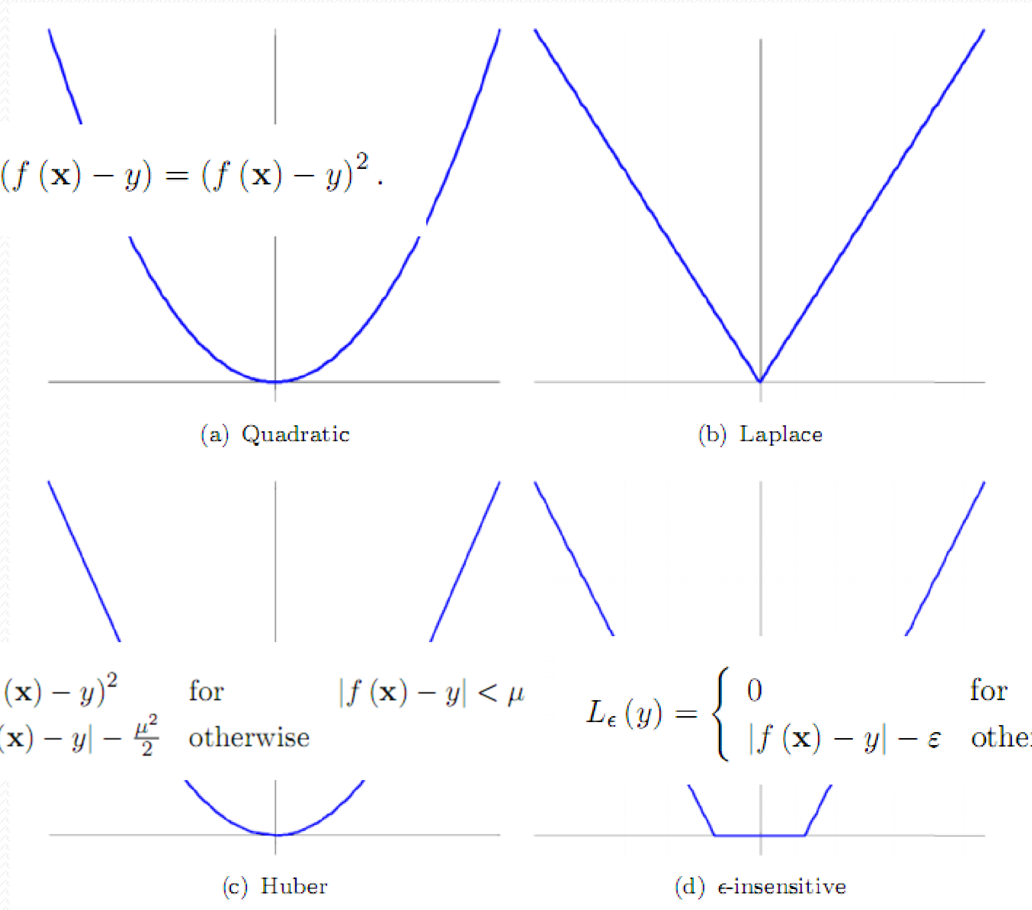


4. SVR- Loss Function

- Hàm lỗi, cho phép các giá trị được phép sai dưới 1 giá trị nhất định.

- Các loại hàm lỗi:
  - a) Quadratic
  - b) Laplace
  - c) Huber
  - d) $\varepsilon -$ insensitive

4. SVR- Loss Function



$$L_{quad}\left(f\left(\mathbf{x}\right)-y\right)=\left(f\left(\mathbf{x}\right)-y\right)^2.$$

(a) Quadratic

(b) Laplace

$$L_{huber}\left(f\left(\mathbf{x}\right)-y\right)=\begin{cases}\frac{1}{2}\left(f\left(\mathbf{x}\right)-y\right)^2 & \text{for} \quad \left|f\left(\mathbf{x}\right)-y\right|<\mu \\ \mu\left|f\left(\mathbf{x}\right)-y\right|-\frac{\mu^2}{2} & \text{otherwise}\end{cases}$$

$$L_{\epsilon}\left(y\right)=\begin{cases}0 & \text{for} \quad \left|f\left(\mathbf{x}\right)-y\right|<\varepsilon \\ \left|f\left(\mathbf{x}\right)-y\right|-\varepsilon & \text{otherwise}\end{cases}.$$

(c) Huber

(d) ϵ-insensitive

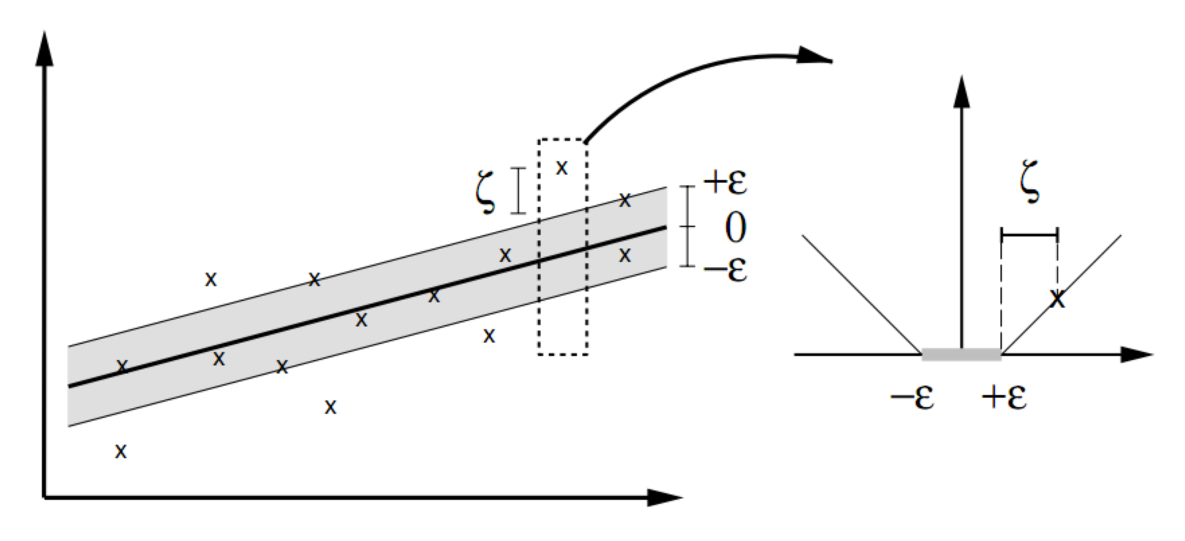
4. SVR- ε – insensitive Loss Function

$$L_\epsilon(y) = \begin{cases} 0 & \text{for} \quad |f(\mathbf{x}) - y| < \varepsilon \\ |f(\mathbf{x}) - y| - \varepsilon & \text{otherwise} \end{cases}$$

- Bài toán lúc này giống bài toán của SVC giải tìm w và b

$$\begin{aligned} \underset{\mathbf{w}\in\mathcal{H},\boldsymbol{\xi}^{(*)}\in\mathbb{R}^m,b\in\mathbb{R}}{\text{minimize}} \quad & \tau(\mathbf{w},\boldsymbol{\xi}^{(*)}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\cdot\frac{1}{m}\sum_{i=1}^{m}(\xi_i+\xi_i^*), \\ \text{subject to } & (\langle\mathbf{w},\mathbf{x}_i\rangle + b) - y_i \le \varepsilon + \xi_i, \\ & y_i - (\langle\mathbf{w},\mathbf{x}_i\rangle + b) \le \varepsilon + \xi_i^*, \\ & \xi_i^{(*)} \ge 0. \end{aligned}$$

4. SVR- ε– insensitive Loss Function



4. SVR- Tính w

- Nhân tử Lagrange

$$L := \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{m}\sum_{i=1}^{m}(\xi_i + \xi_i^*) - \sum_{i=1}^{m}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$
$$- \sum_{i=1}^{m}\alpha_i(\varepsilon + \xi_i + y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - b)$$
$$- \sum_{i=1}^{m}\alpha_i^*(\varepsilon + \xi_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i\rangle + b),$$

- Với điều kiện

$$\alpha_i^{(*)}, \eta_i^{(*)} \geq 0.$$

4. SVR- Tính w

$$
\begin{aligned}
\partial_b L &= \quad \textstyle\sum_{i=1}^{m}(\alpha_i - \alpha_i^*) &&= 0,\\
\partial_{\mathbf{w}} L &= \mathbf{w} - \textstyle\sum_{i=1}^{m}(\alpha_i^* - \alpha_i)\mathbf{x}_i &&= 0,\\
\partial_{\xi_i^{(*)}} L &= \quad \frac{C}{m} - \alpha_i^{(*)} - \eta_i^{(*)} &&= 0.
\end{aligned}
$$

$$
\begin{aligned}
\underset{\boldsymbol{\alpha}^{(*)} \in \mathbb{R}^m}{\text{maximize}} \quad & \begin{cases} -\frac{1}{2} \sum\limits_{i,j=1}^{m} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ -\varepsilon \sum\limits_{i=1}^{m} (\alpha_i^* + \alpha_i) + \sum\limits_{i=1}^{m} y_i(\alpha_i^* - \alpha_i), \end{cases} \\
\text{subject to} \quad & \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C/m].
\end{aligned}
$$

4. SVR- Tính w

- w tính theo công thức

$$\mathbf{w} = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i)\mathbf{x}_i, \quad \text{thus} \quad f(\mathbf{x}) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) \langle \mathbf{x}_i, \mathbf{x} \rangle + b.$$

4. SVR- Tính b

- Dùng điều kiện Karush-Kuhn-Tucker (KTT)

$$\begin{aligned}
\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) &= 0 \\
\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) &= 0 \\
(C - \alpha_i)\xi_i &= 0 \\
(C - \alpha_i^*)\xi_i^* &= 0.
\end{aligned}$$

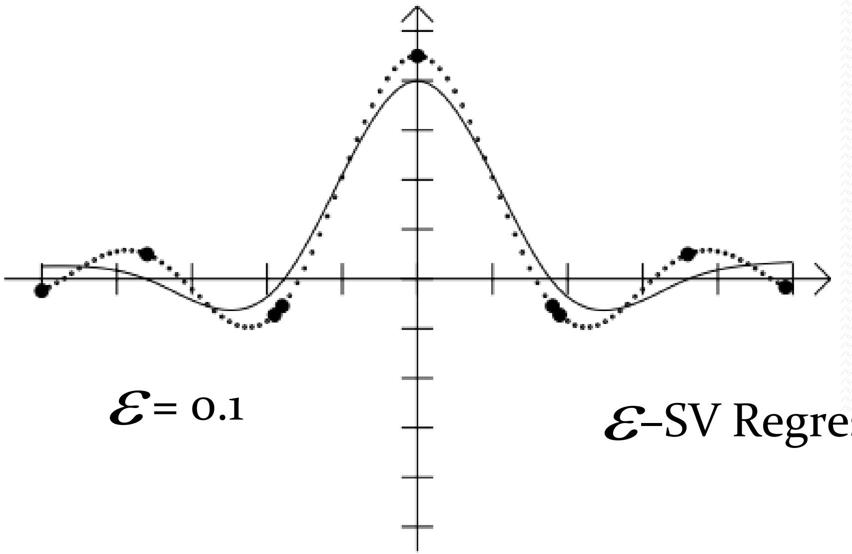
$$\begin{array}{lllll}
\varepsilon - y_i + \langle w, x_i \rangle + b \geq 0 & \text{and} & \xi_i = 0 & \text{if} & \alpha_i < C \\
\varepsilon - y_i + \langle w, x_i \rangle + b \leq 0 & & & \text{if} & \alpha_i > 0
\end{array}$$

4. SVR- Tính b

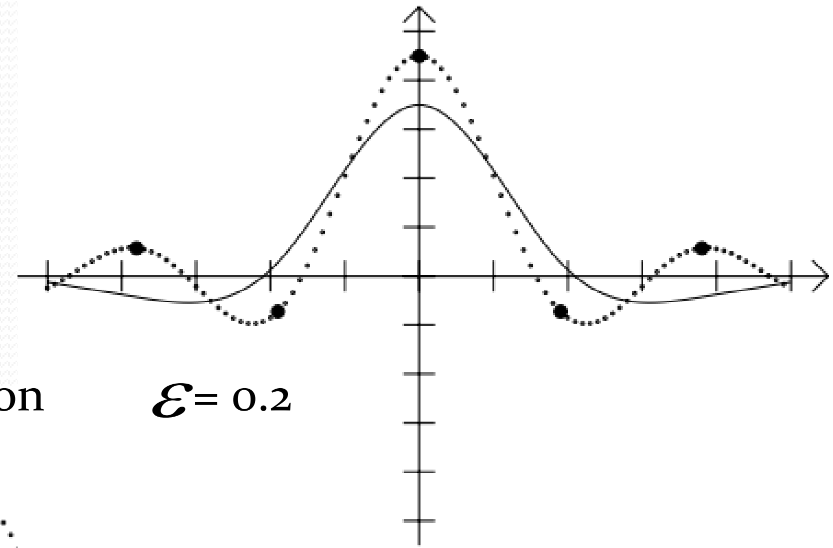
- b được tính theo công thức

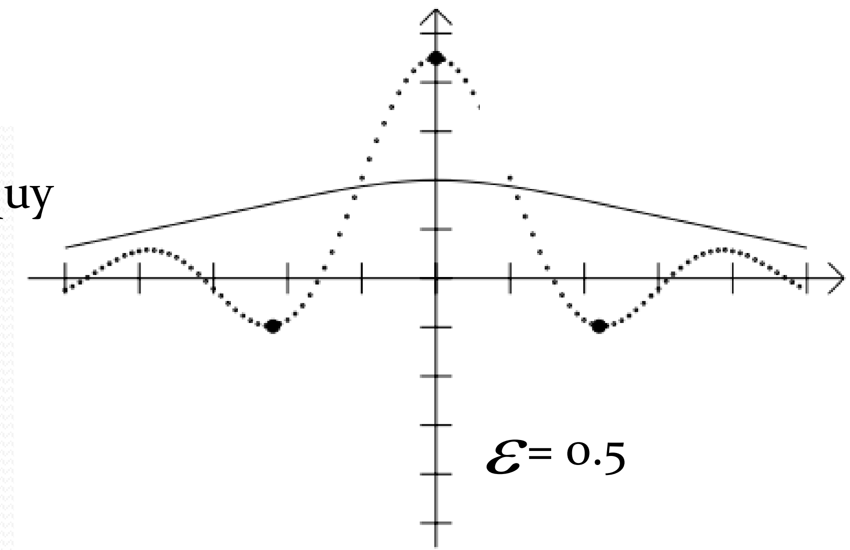
$$\max\{-\varepsilon + y_i - \langle w, x_i\rangle | \alpha_i < C \text{ or } \alpha_i^* > 0\} \le b \le$$
$$\min\{-\varepsilon + y_i - \langle w, x_i\rangle | \alpha_i > 0 \text{ or } \alpha_i^* < C\}$$

4. SVR- Nhận xét



ε–SV Regression





Đường liền nét: hàm hồi quy
Đường chấm: dữ liệu
Chấm đậm: SV

4. SVR- *ν-SV Regression*

- Chỉnh sửa từ ε–SV Regression
  - ε được xác định trong quá trình tính toán
  - ν - là hằng số không âm.

$$\underset{\mathbf{w}\in\mathcal{H},\boldsymbol{\xi}^{(*)}\in\mathbb{R}^m,\varepsilon,b\in\mathbb{R}}{\text{minimize}} \quad \tau(\mathbf{w},\boldsymbol{\xi}^{(*)},\varepsilon) = \frac{1}{2}\|\mathbf{w}\|^2 + C\cdot\left(\nu\varepsilon + \frac{1}{m}\sum_{i=1}^{m}(\xi_i+\xi_i^*)\right),$$

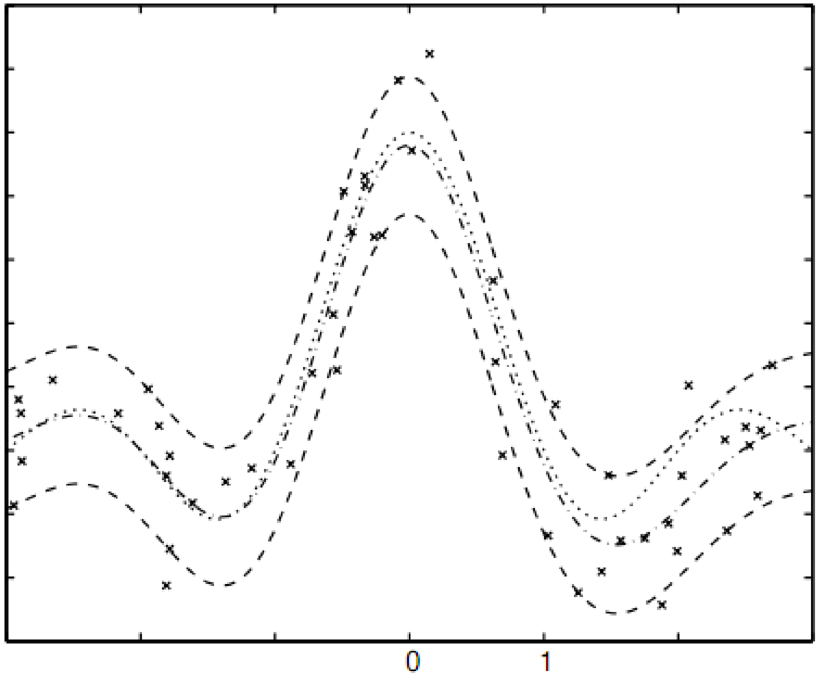
$$\begin{aligned}\text{subject to } & (\langle\mathbf{w},\mathbf{x}_i\rangle + b) - y_i \le \varepsilon + \xi_i,\\ & y_i - (\langle\mathbf{w},\mathbf{x}_i\rangle + b) \le \varepsilon + \xi_i^*,\\ & \xi_i^{(*)} \ge 0,\ \ \varepsilon \ge 0.\end{aligned}$$

4. SVR- *ν là bao nhiêu?*

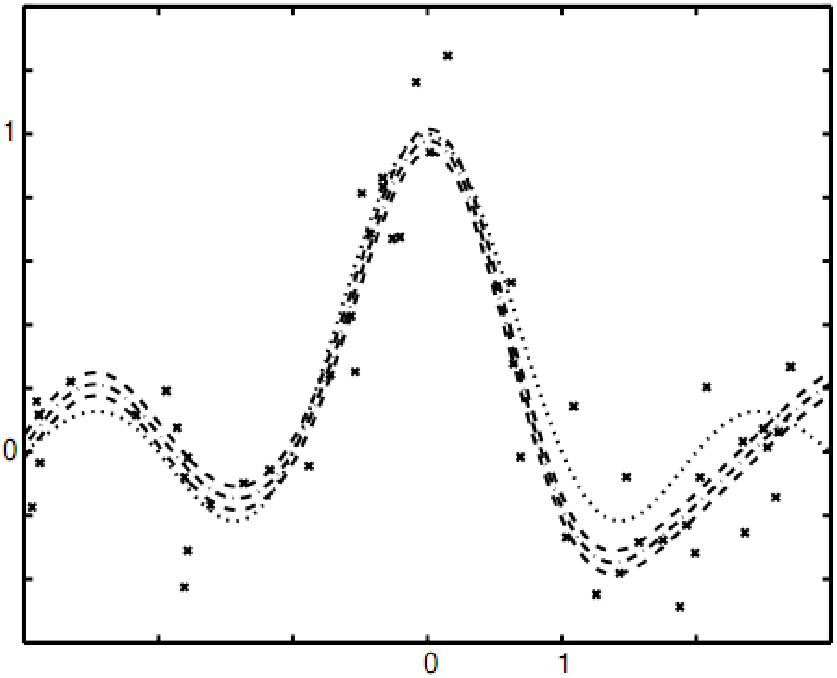
- *ν – lớn hơn tỉ lệ lỗi (fraction of errors)*
- *ν – nhỏ hơn tỉ lệ SV(fraction of SVs)*
- *Khi số mẫu tăng thì ν sẽ tiến về 1 giá trị*

m	10	50	100	200	500	1000	1500	2000
ε	0.27	0.22	0.23	0.25	0.26	0.26	0.26	0.26
fraction of errors	0.00	0.10	0.14	0.18	0.19	0.20	0.20	0.20
fraction of SVs	0.40	0.28	0.24	0.23	0.21	0.21	0.20	0.20

4. SVR- *ν-SV Regression*



$\varepsilon = 0.22$
$\nu = 0.2$

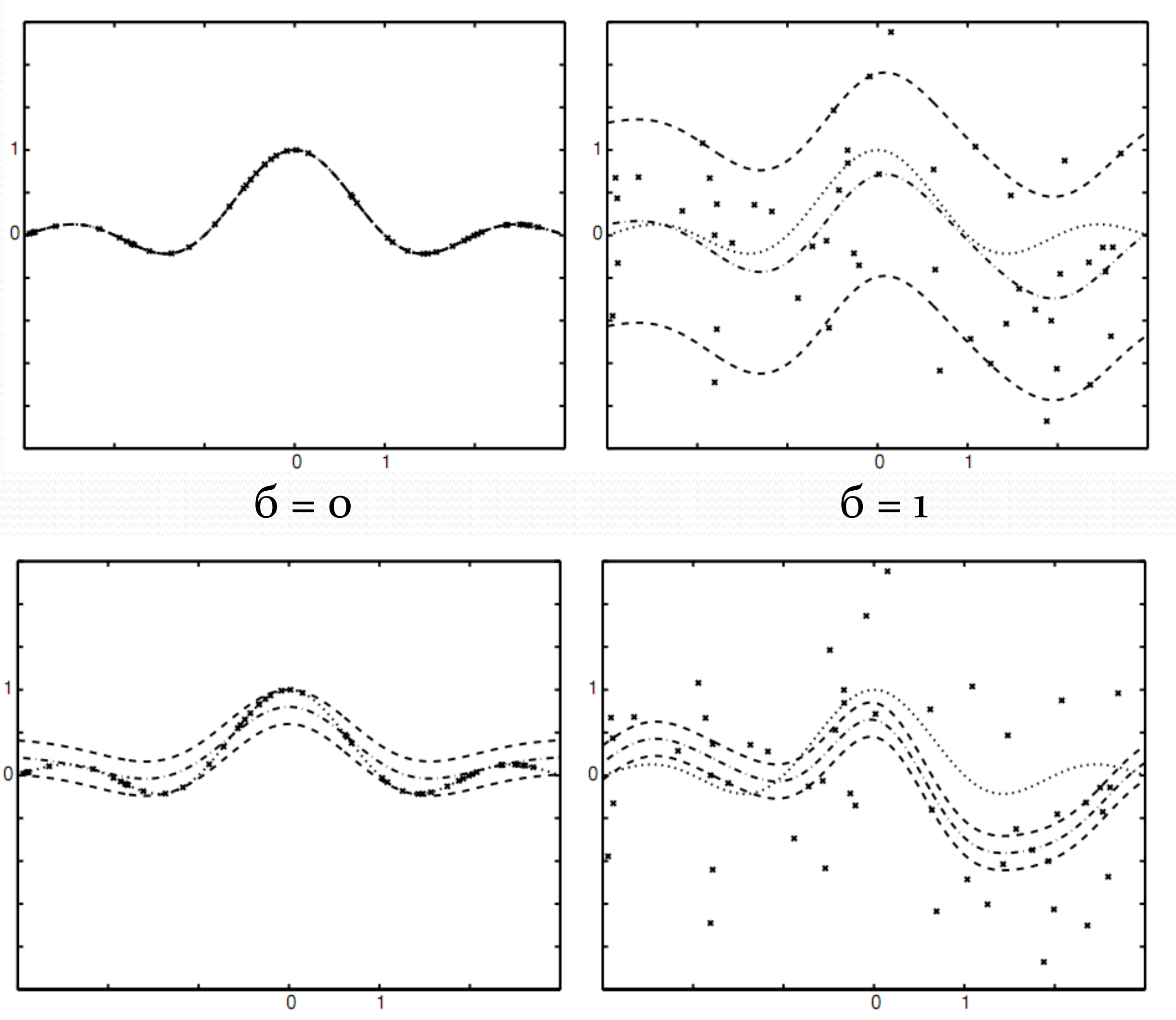


$\varepsilon = 0.04$
$\nu = 0.8$

4. SVR- So sánh

$\varepsilon = 0$
$\nu = 0.2$

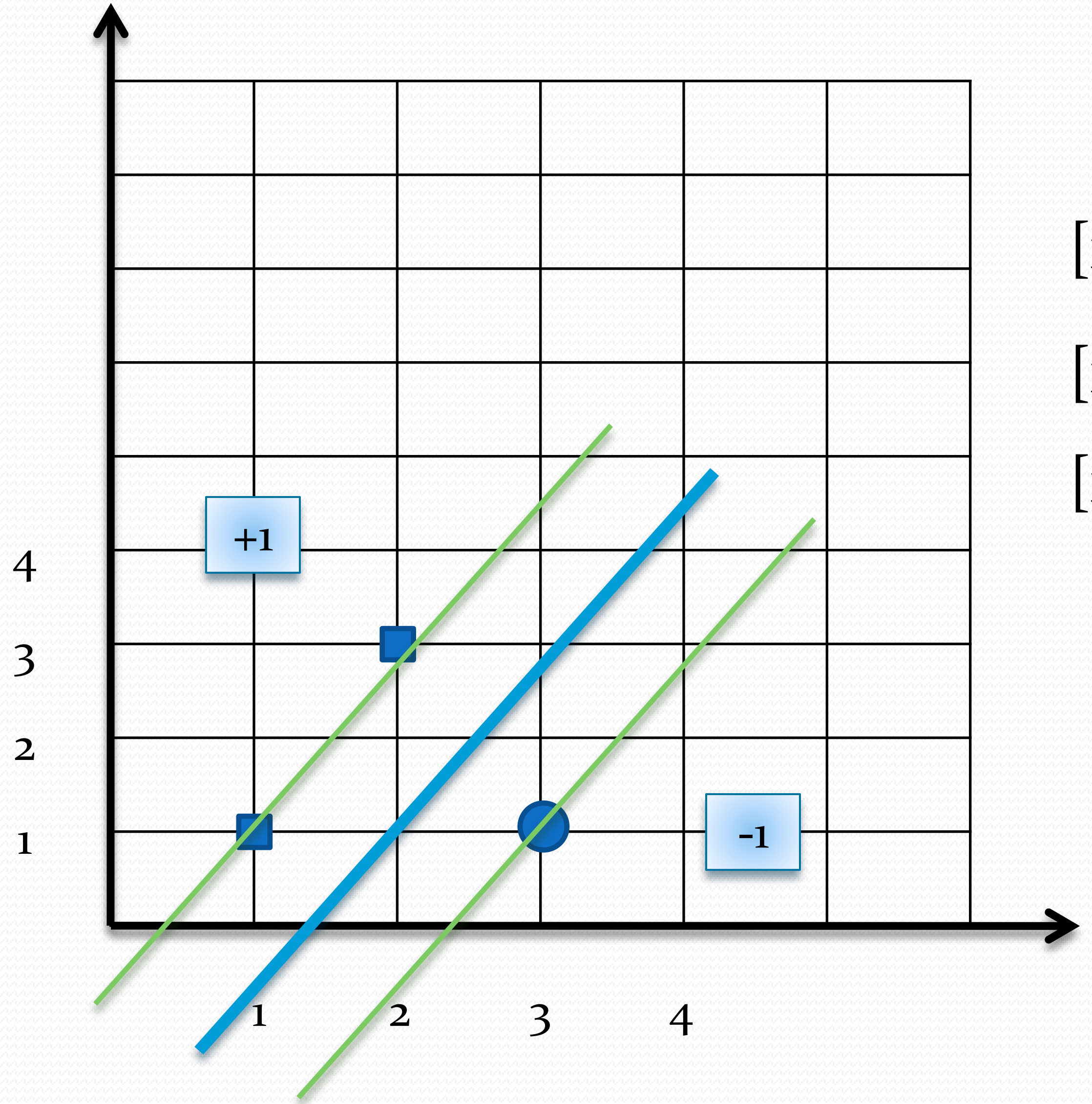
$\varepsilon = 1.19$
$\nu = 0.2$



$\varepsilon = 0.2$

$\varepsilon = 0.2$

5. Ví dụ



[x1(1,1), 1]

[x2(2,3), 1]

[x3(3,1), -1]

5. Ví dụ

- {[x1(1,1),1] ; [x2(2,3),1] ; [x3(3,1),-1]}
- Để giải bài toán đối ngẫu, ta tìm cực tiểu của:

$$L(w,b,\alpha) = \frac{1}{2}w^T w - \sum_{i=1}^{n} \alpha_i \left[y_i (w^T x_i + b) - 1 \right]$$

- Cực tiểu của L xảy ra tại w và b sao cho:

$$\frac{\partial L(w,b,\alpha)}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \Rightarrow \boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

$$w = \alpha_1(1,1) + \alpha_2(2,3) - \alpha_3(3,1)$$
$$= (\alpha_1 + 2\alpha_2 - 3\alpha_3, \alpha_1 + 3\alpha_2 - \alpha_3)$$

5. Ví dụ

- Cực tiểu của L xảy ra tại w và b sao cho:

$$w = (\alpha_1 + 2\alpha_2 - 3\alpha_3, \alpha_1 + 3\alpha_2 - \alpha_3)$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = \alpha_1 + \alpha_2 - \alpha_3 = 0 \qquad (1)$$

- Mặt khác, sử dụng điều kiện KKT:

$$\alpha_i\left[y_i(w^T x_i + b) - 1\right] = 0; i = 1,\ldots,n$$

$$\begin{cases} \alpha_1[w(1,1) + b - 1] = 0 & (2) \\ \alpha_2[w(2,3) + b - 1] = 0 & (3) \\ \alpha_3[-w(3,1) - b - 1] = 0 & (4) \end{cases}$$

[x1(1,1), 1]

[x2(2,3), 1]

[x3(3,1), -1]

5. Ví dụ

- Điều kiện KKT

$$\begin{cases} \alpha_1[w(1,1)+b-1]=0 \\ \alpha_2[w(2,3)+b-1]=0 \\ \alpha_3[-w(3,1)-b-1]=0 \end{cases}$$

$$w=(\alpha_1+2\alpha_2-3\alpha_3,\alpha_1+3\alpha_2-\alpha_3)$$

$$\begin{cases} \alpha_1(2\alpha_1+5\alpha_2-4\alpha_3+b-1)=0 \\ \alpha_2(5\alpha_1+13\alpha_2-9\alpha_3+b-1)=0 \\ \alpha_3(4\alpha_1+9\alpha_2-10\alpha_3+b+1]=0 \end{cases}$$

5. Ví dụ

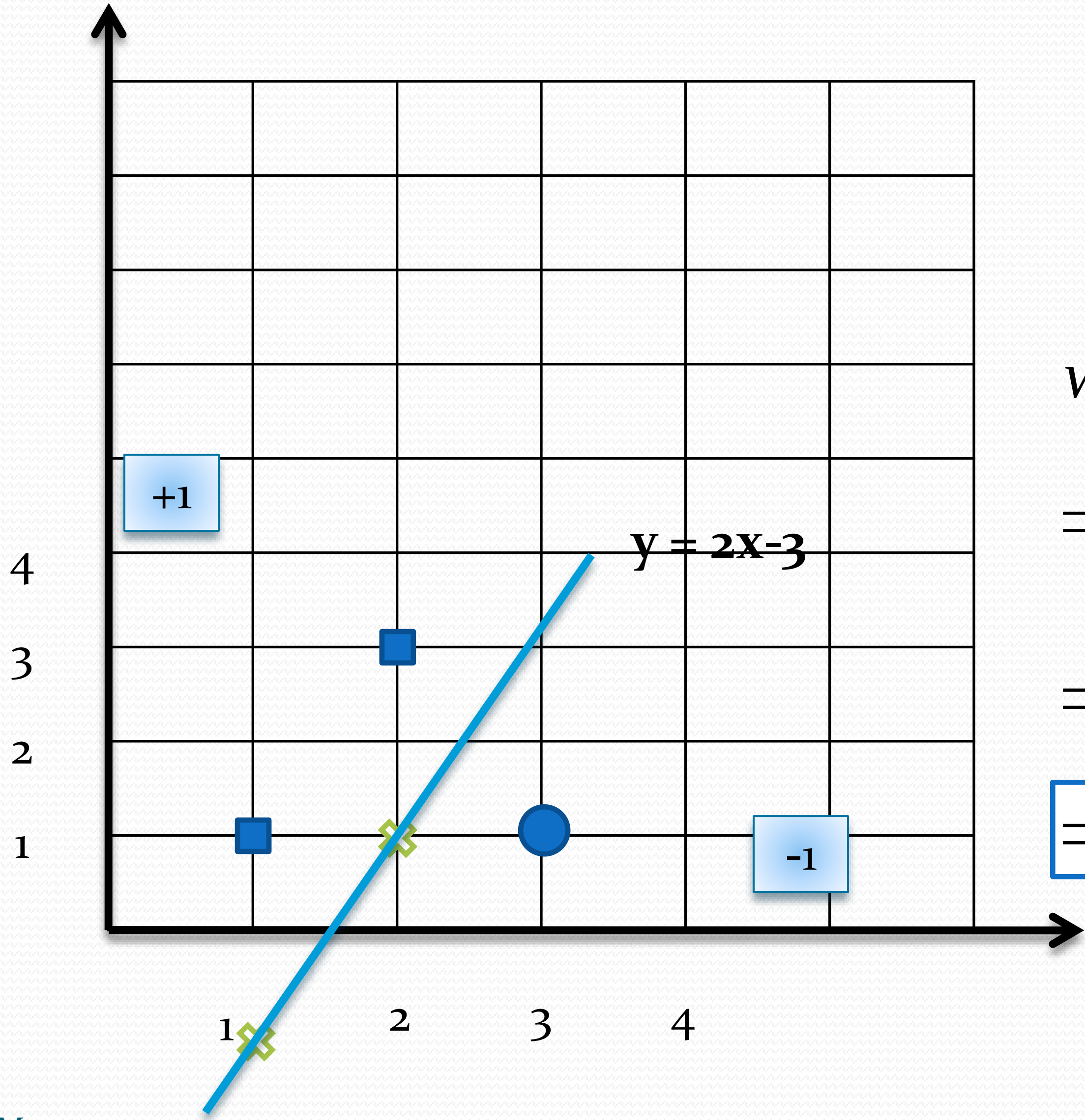
$$\begin{cases} \alpha_1(2\alpha_1 + 5\alpha_2 - 4\alpha_3 + b - 1) = 0 \\ \alpha_2(5\alpha_1 + 13\alpha_2 - 9\alpha_3 + b - 1) = 0 \\ \alpha_3(4\alpha_1 + 9\alpha_2 - 10\alpha_3 + b + 1] = 0 \\ \alpha_1 + \alpha_2 - \alpha_3 = 0 \end{cases}$$

- Giải hệ 4 phương trình 4 ẩn ta được

$$\begin{cases} \alpha_1 = 3/8 \\ \alpha_2 = 1/4 \\ \alpha_3 = 5/8 \\ b = 3/2 \end{cases}$$

$$\boxed{\begin{cases} w = (-1, \frac{1}{2}) \\ b = 3/2 \end{cases}}$$

5. Ví dụ



$$\begin{cases} w = (-1, \frac{1}{2}) \\ b = 3/2 \end{cases}$$

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

$$\Rightarrow (-1, \frac{1}{2})(x, y) + \frac{3}{2} = 0$$

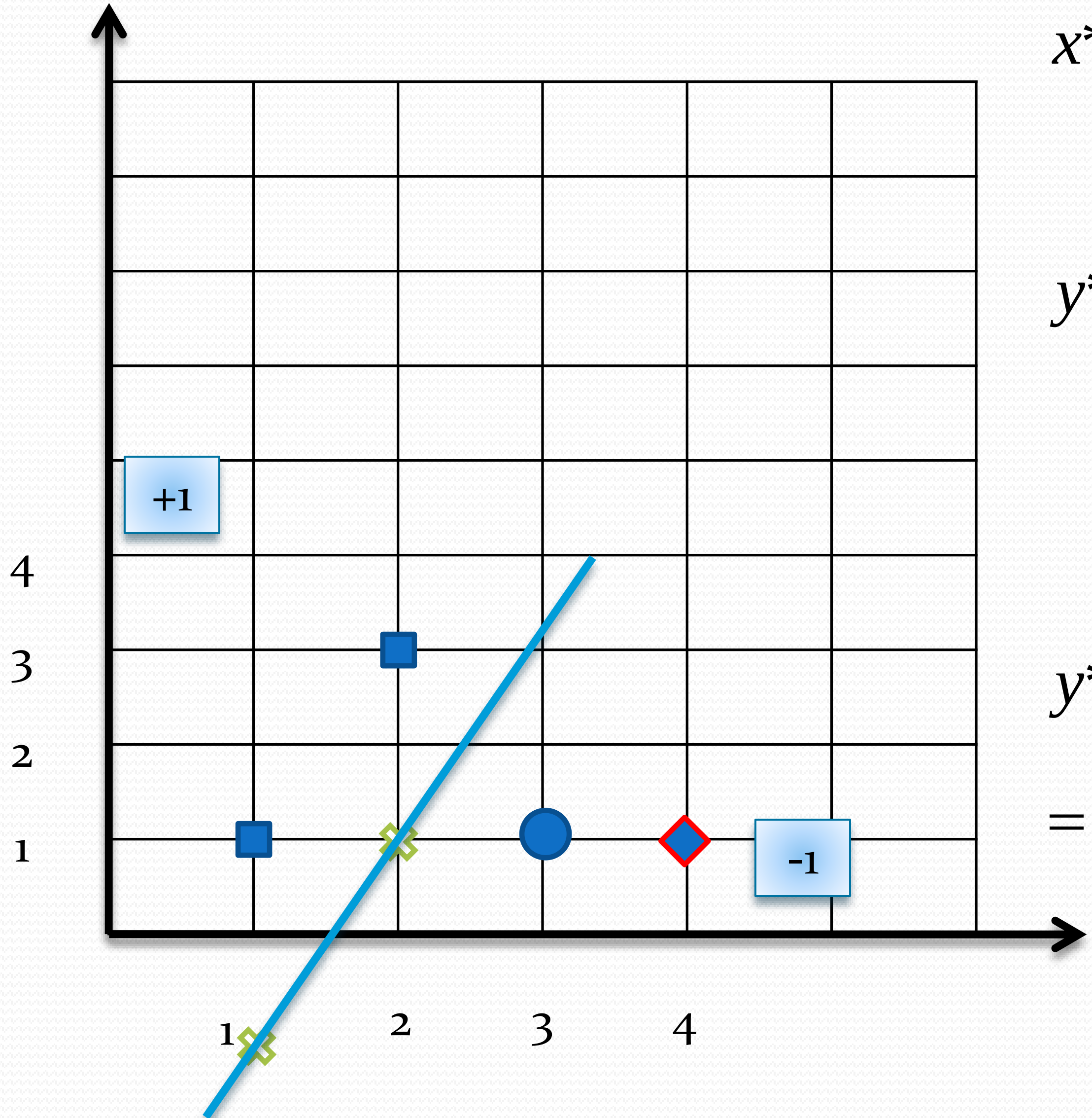
$$\Rightarrow -x + \frac{1}{2} y = -\frac{3}{2}$$

$$\boxed{\Rightarrow y = 2x - 3}$$

$$x = 1 \Rightarrow y = -1$$

$$x = 2 \Rightarrow y = 1$$

5. Ví dụ - Testing

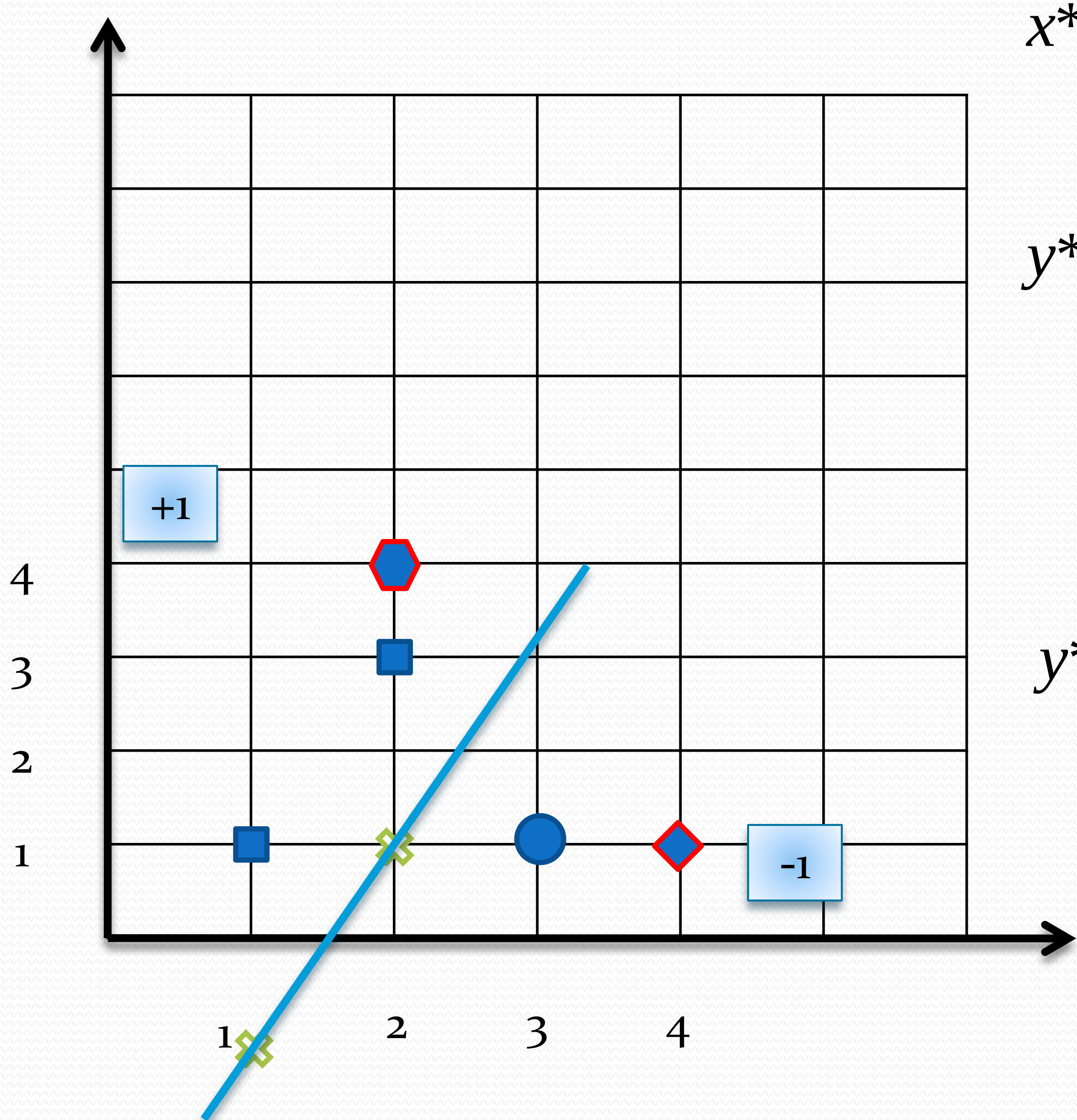


$$x^* = (4,1)$$

$$y^* = sign(w^T x^* + b)$$
$$= sign(-1x + \frac{1}{2}y + \frac{3}{2})$$

$$y^* = sign(-4 + \frac{1}{2} + \frac{3}{2})$$
$$= sign(-2) \Rightarrow y^* = -1$$

5. Ví dụ - Testing



$$x^* = (2,4)$$

$$y^* = sign(w^T x^* + b)$$

$$= sign(-1x + \frac{1}{2} y + \frac{3}{2})$$

$$y^* = sign(-2 + 2 + \frac{3}{2})$$

$$= sign(\frac{3}{2}) \Rightarrow y^* = 1$$