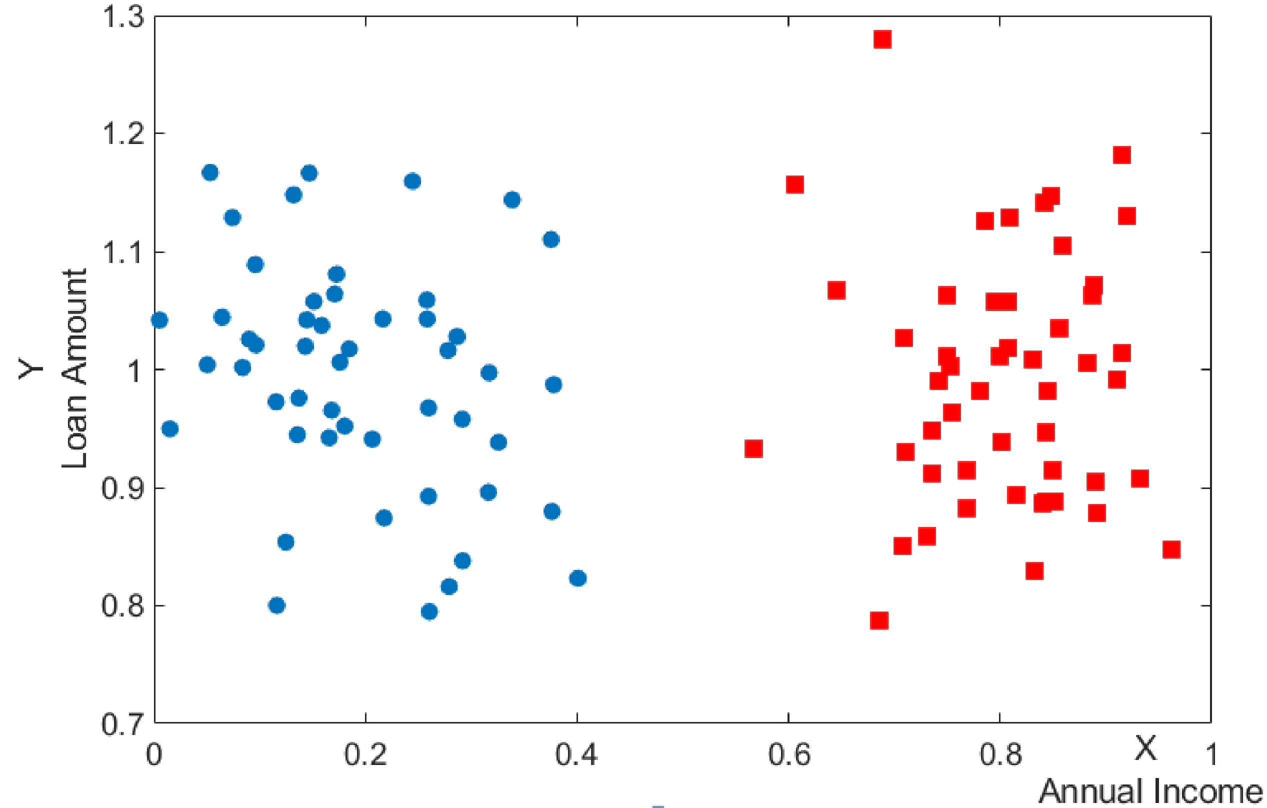
Support Vector Machines

Dr. Pritam Anand.
Assistant Professor,
DA-IICT, Gandhinagar.

Load Defaulter Dataset

Index	Employed	Bank Balance (in thousands rupees)	Annual Salary (in million rupees)	Loan Amount (in thousands rupees)	Defaulted
1	1	0.4721	0.1358	0.9448	0
2	0	0.8412	0.3169	0.9972	0
3	0	0.3687	0.1249	0.8539	0
4	0	0.2547	0.8416	1.1406	1
5	1	0.3111	0.7502	1.014	1

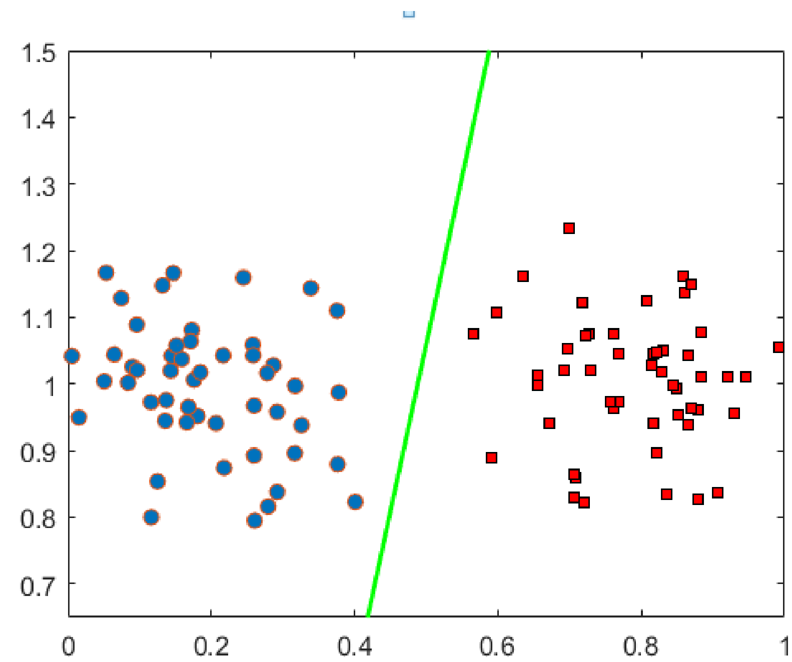
Logistic Regression Decision Rule



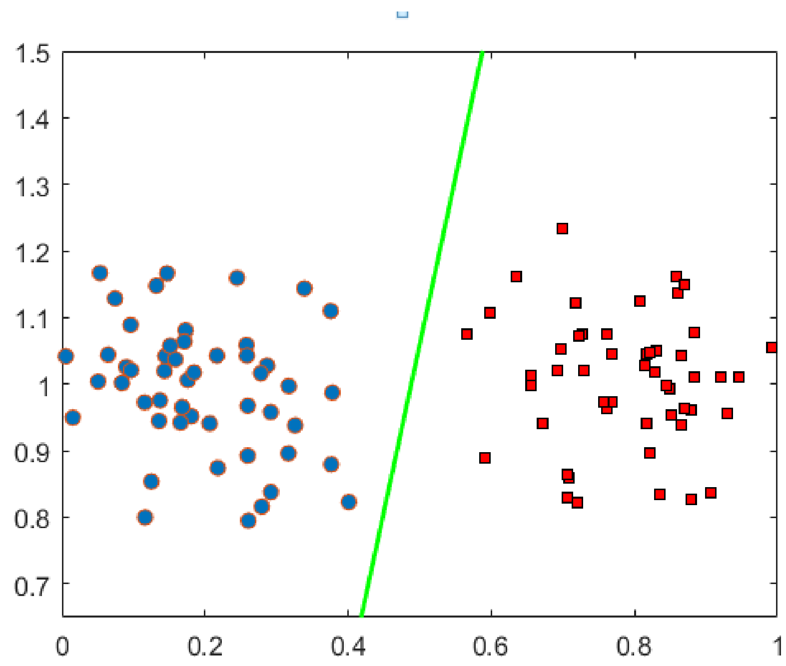
Logistic Regression Classifier

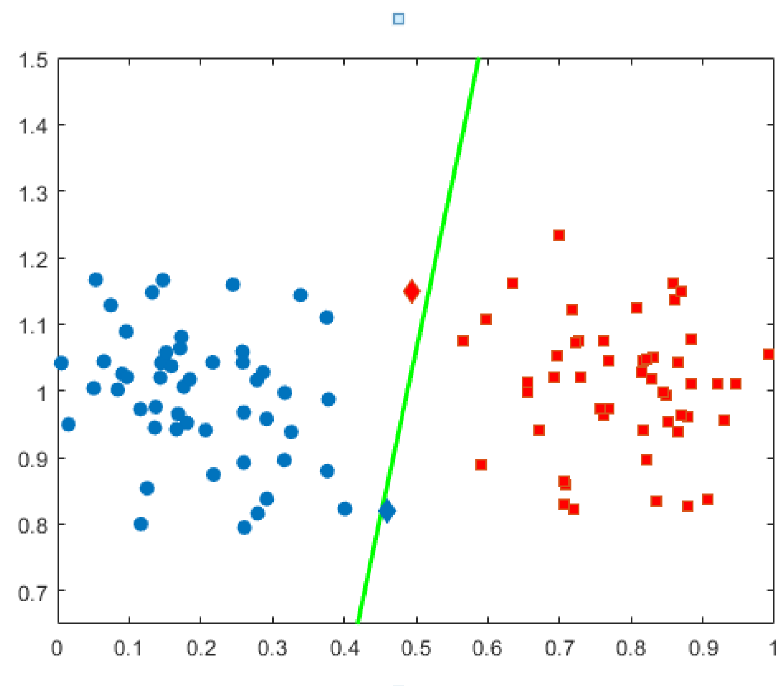
$\beta_1 x_1 + \beta_2 x_2 + \beta_0 = 0$

What is Problem ??

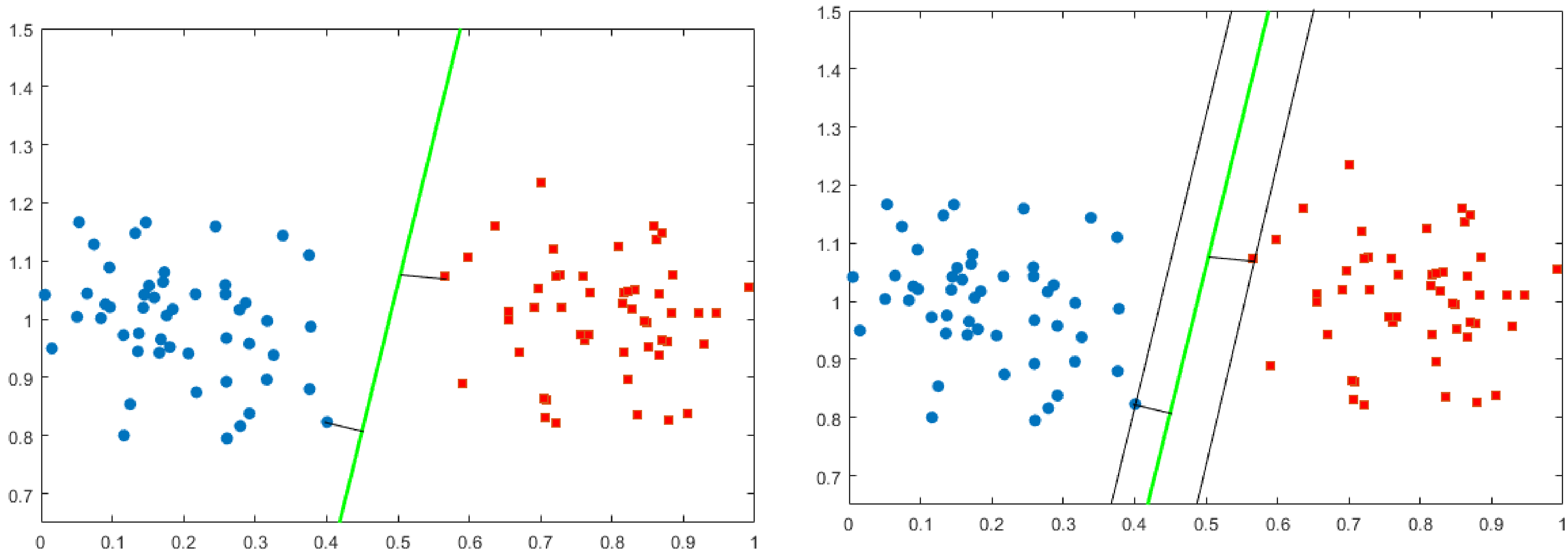


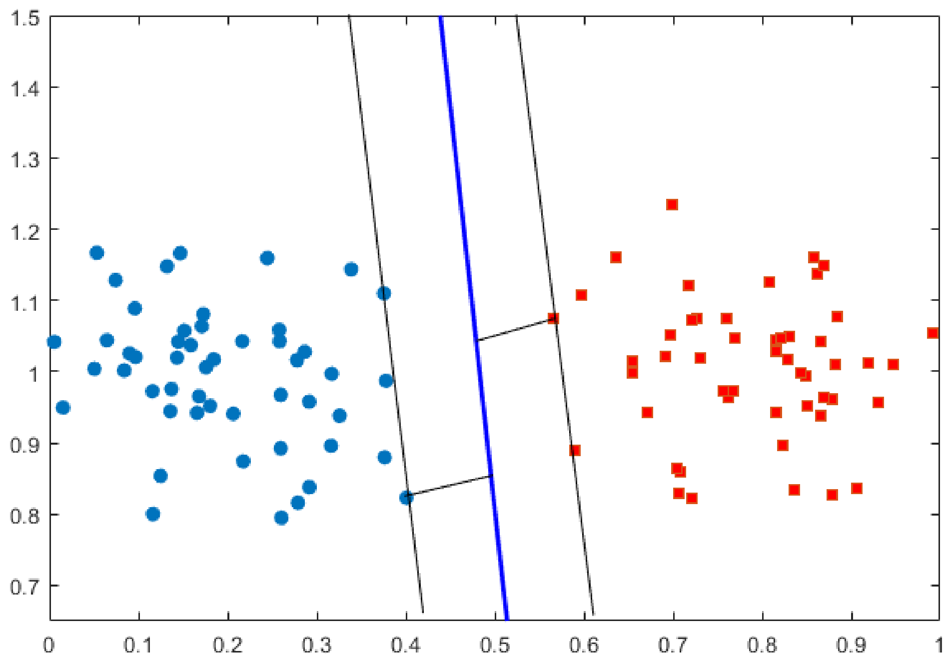
More Susceptible to misclassification



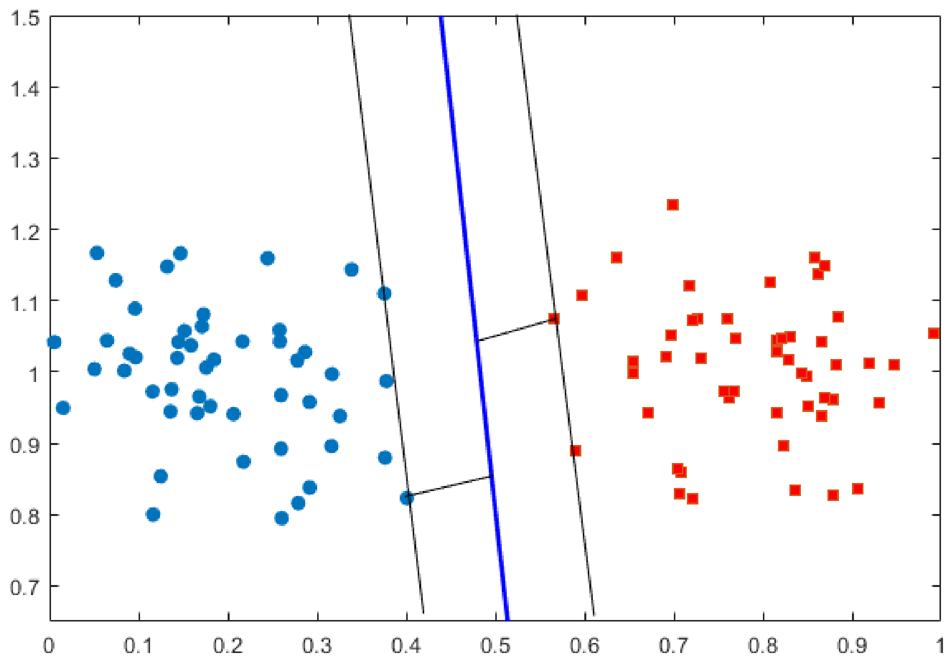


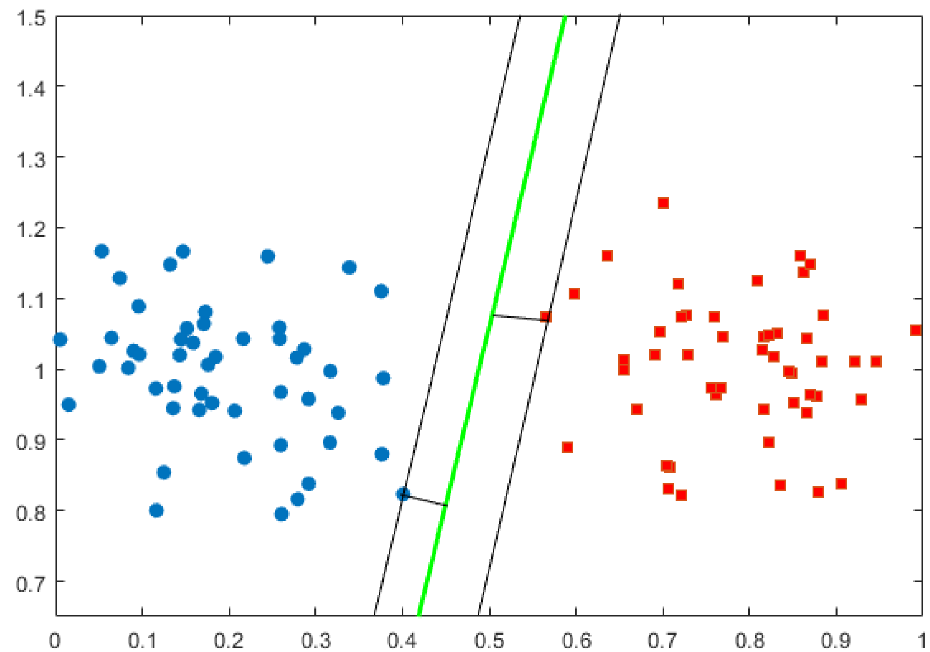
Margin of Logistic Decision function



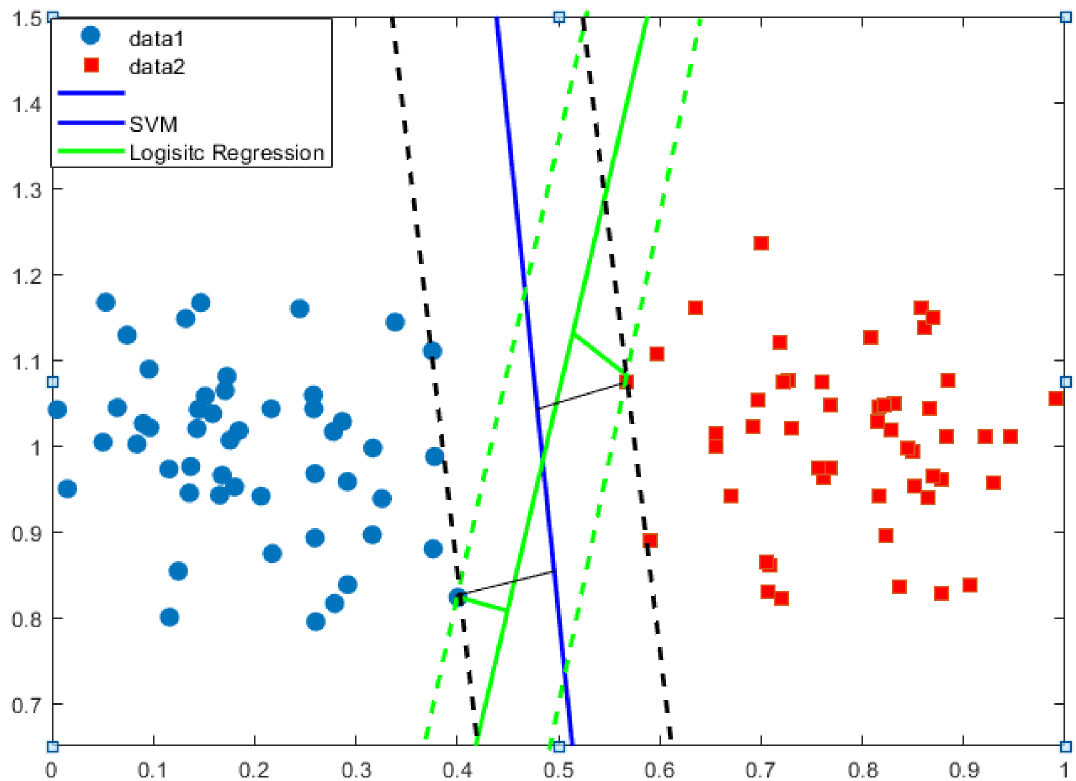

1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Maximal Margin Classifier

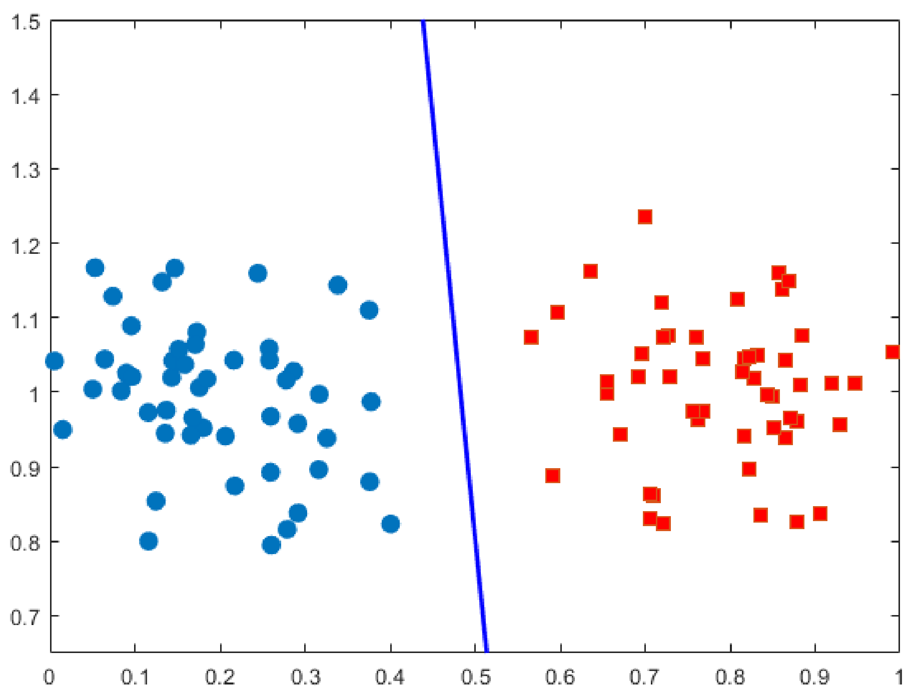


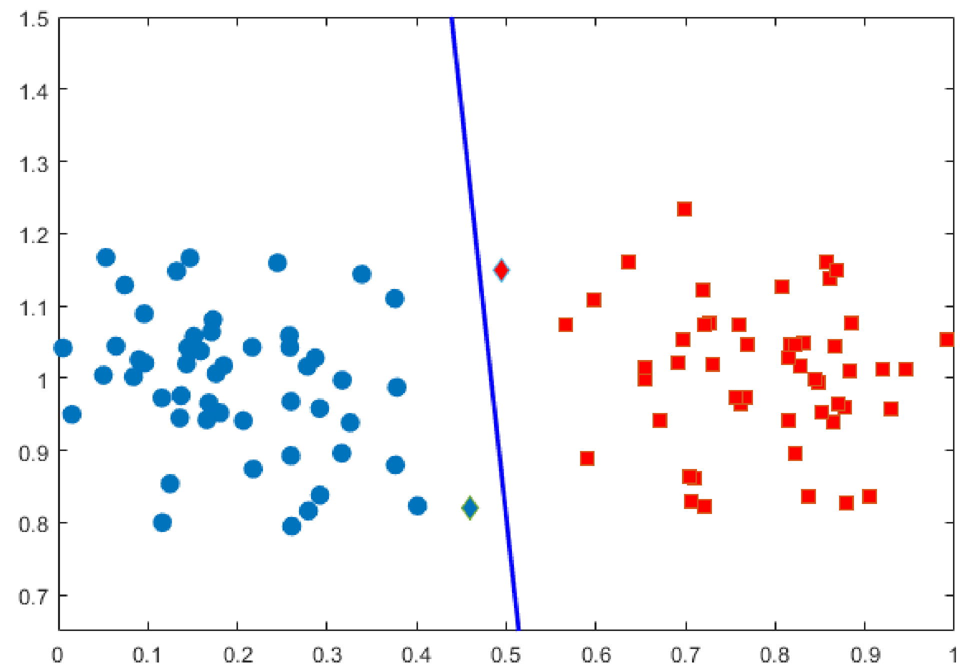


Width of margin

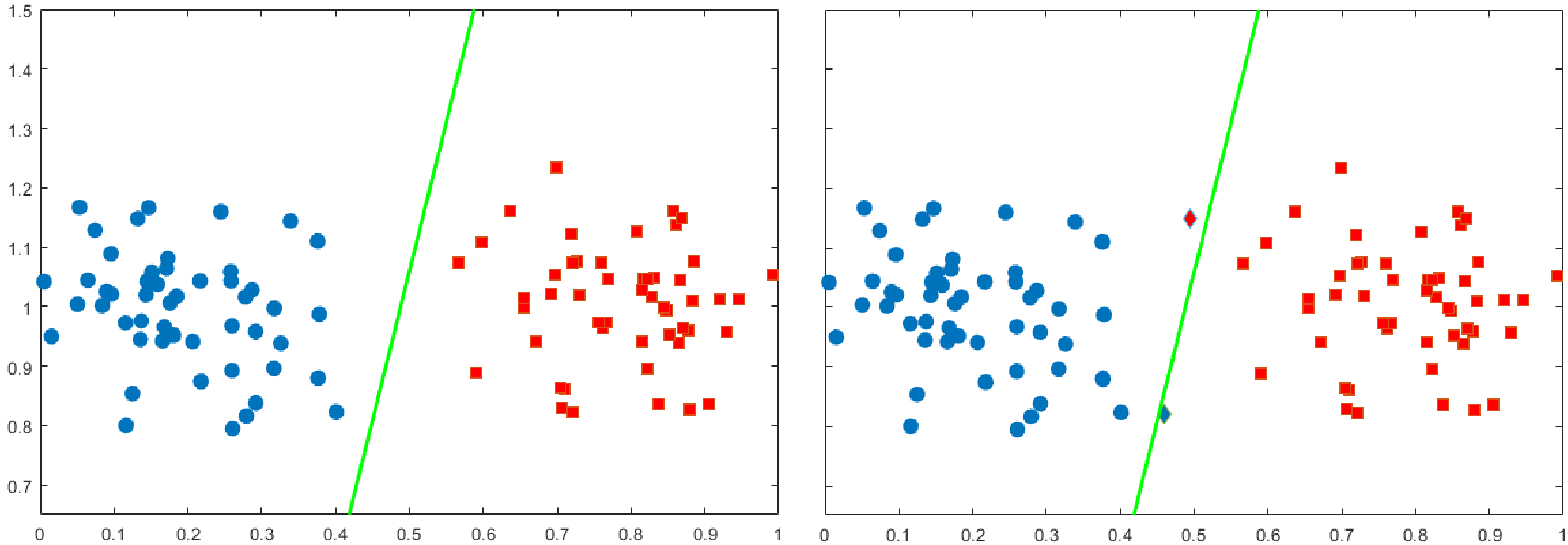


Checking the SVM

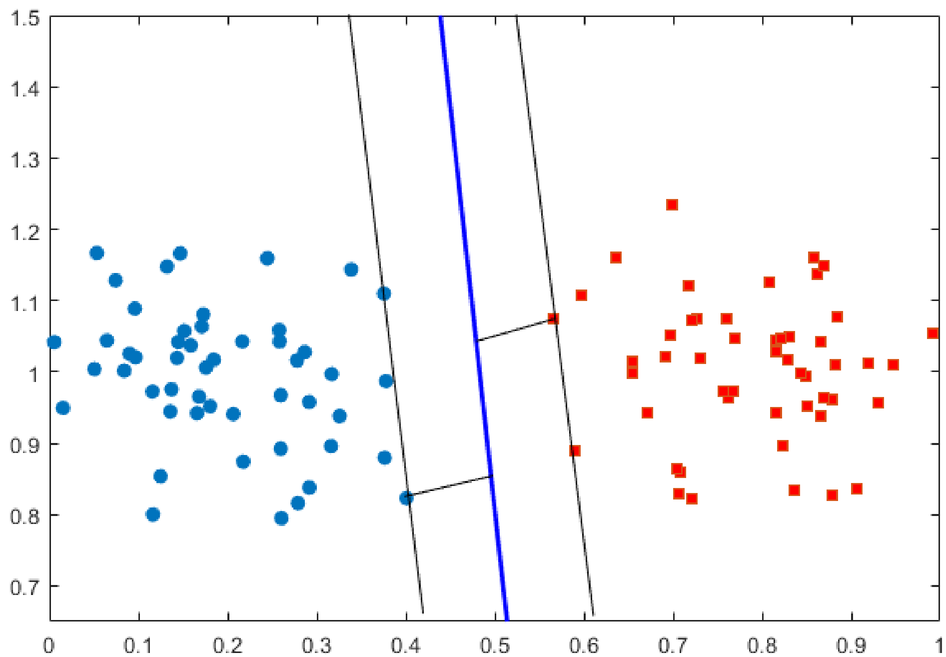




Checking the Logistic Decision boundary



Support Vectors

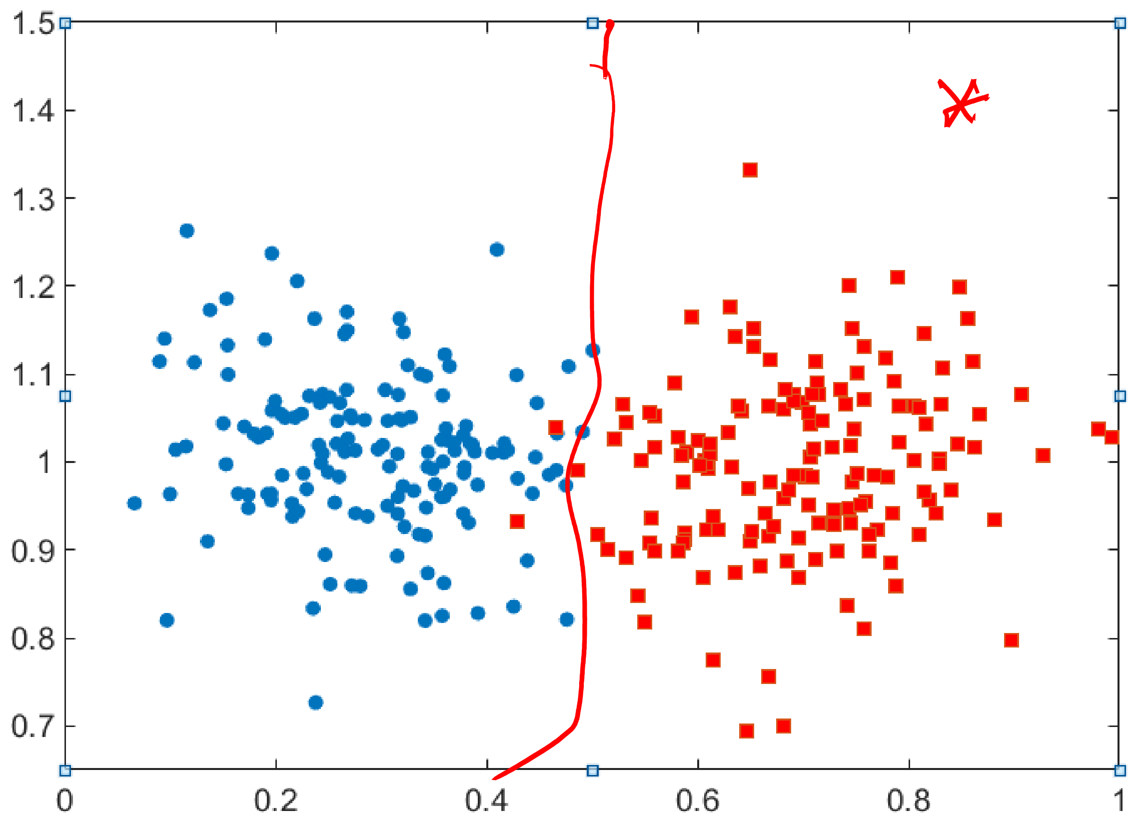


What to do for this

$f(x)$

Si

$\underline{\text{Sign}(f(x_{test}))}$



$\sqrt{w_1^2 + w_2^2 \cdots w_n^2}$

$X_+ = \arg\min_{x \in I_2} \frac{w^T x + b}{\|w\|}$

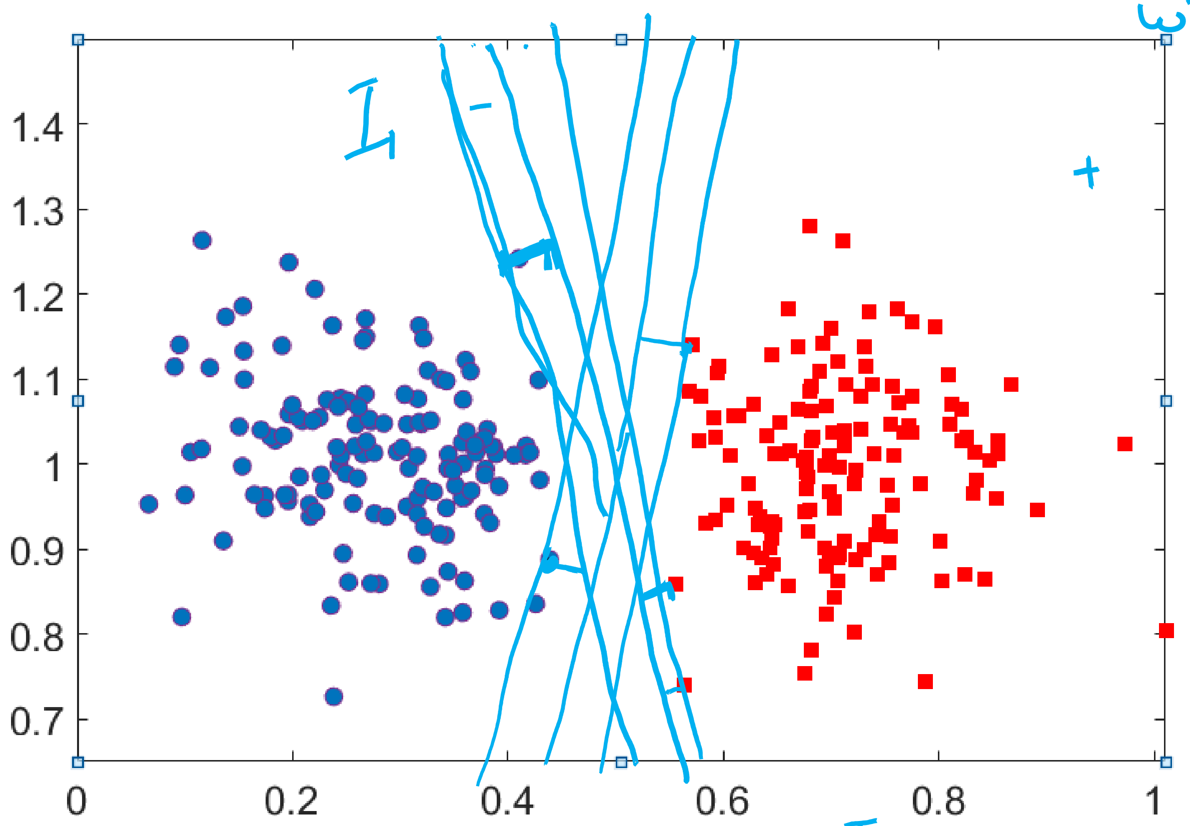
$\max \frac{2\delta}{\|w\|}$

$w^T x + b = \delta$

$w^T x + b = -\delta$

$\left(\frac{w_1}{\delta}\right)^2 +$

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



$w^T x + b = 0$

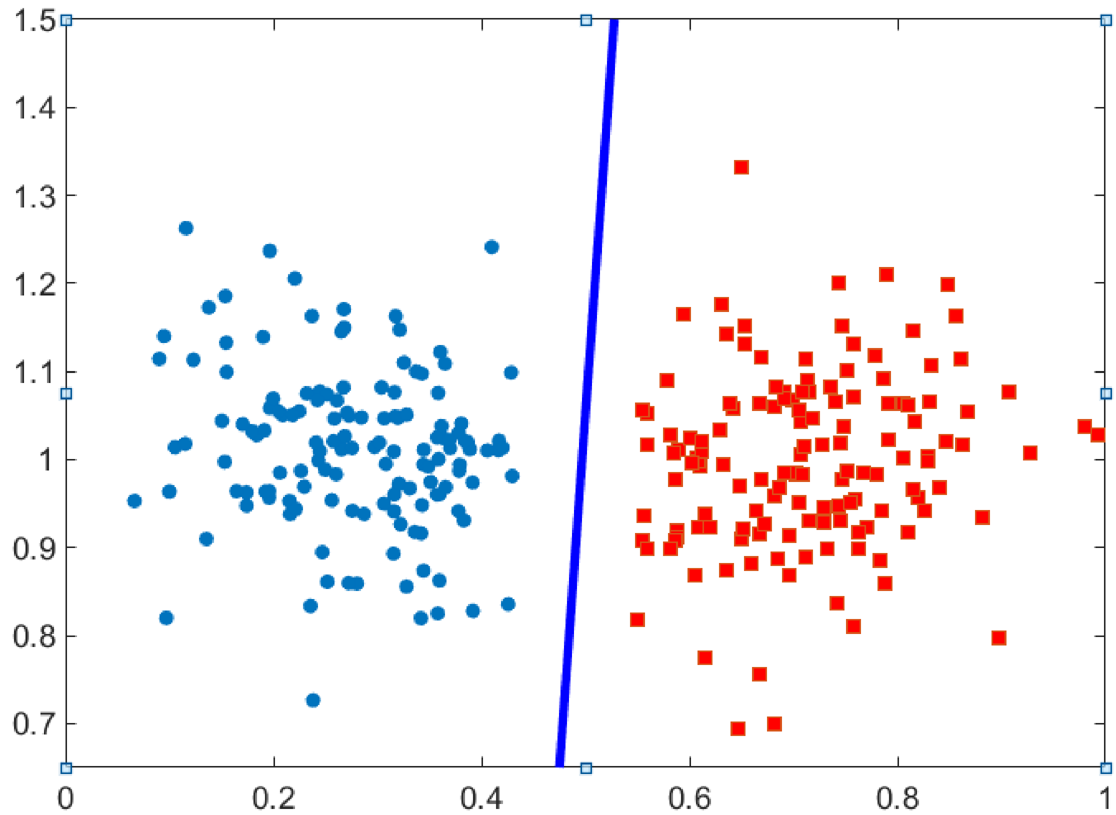
$\frac{w^T}{\delta} x + \frac{b}{\delta} = 0$

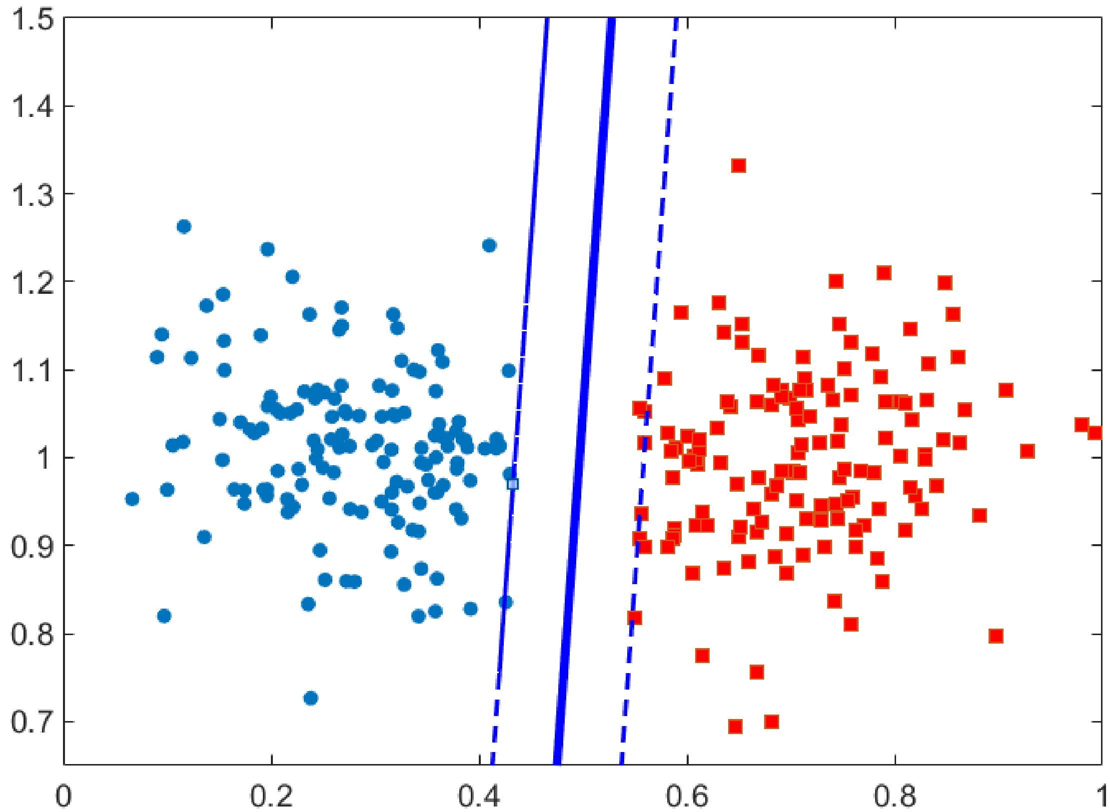
$\max \frac{2}{\|\bar{w}\|}$

$\bar{w}^T x + \bar{b} = -1$

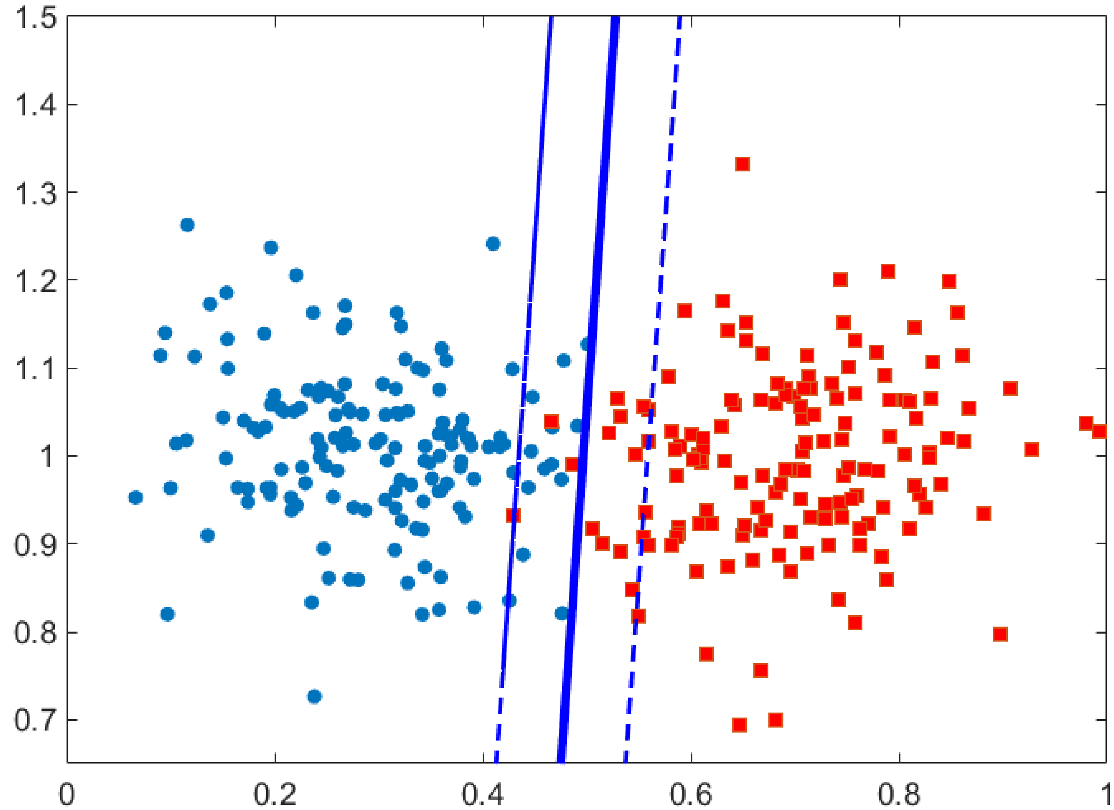
$\bar{w}^T x + \bar{b} = 1$

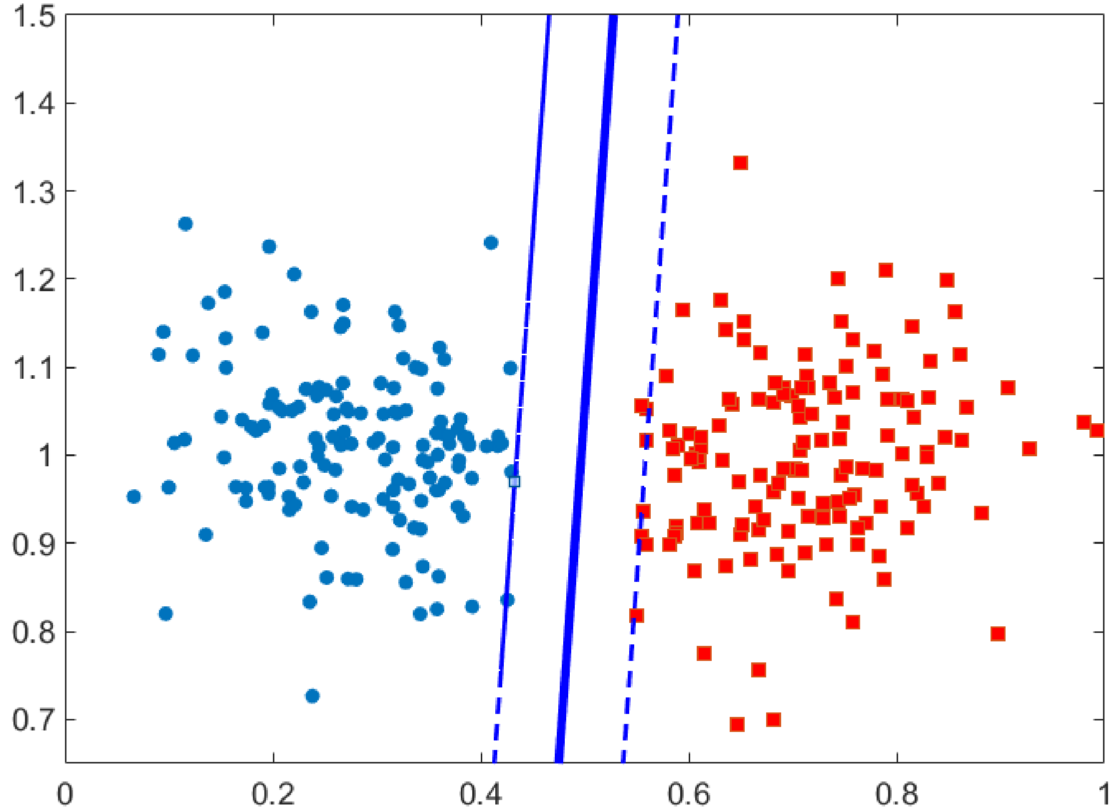
$$\max \frac{2}{\|w\|}$$



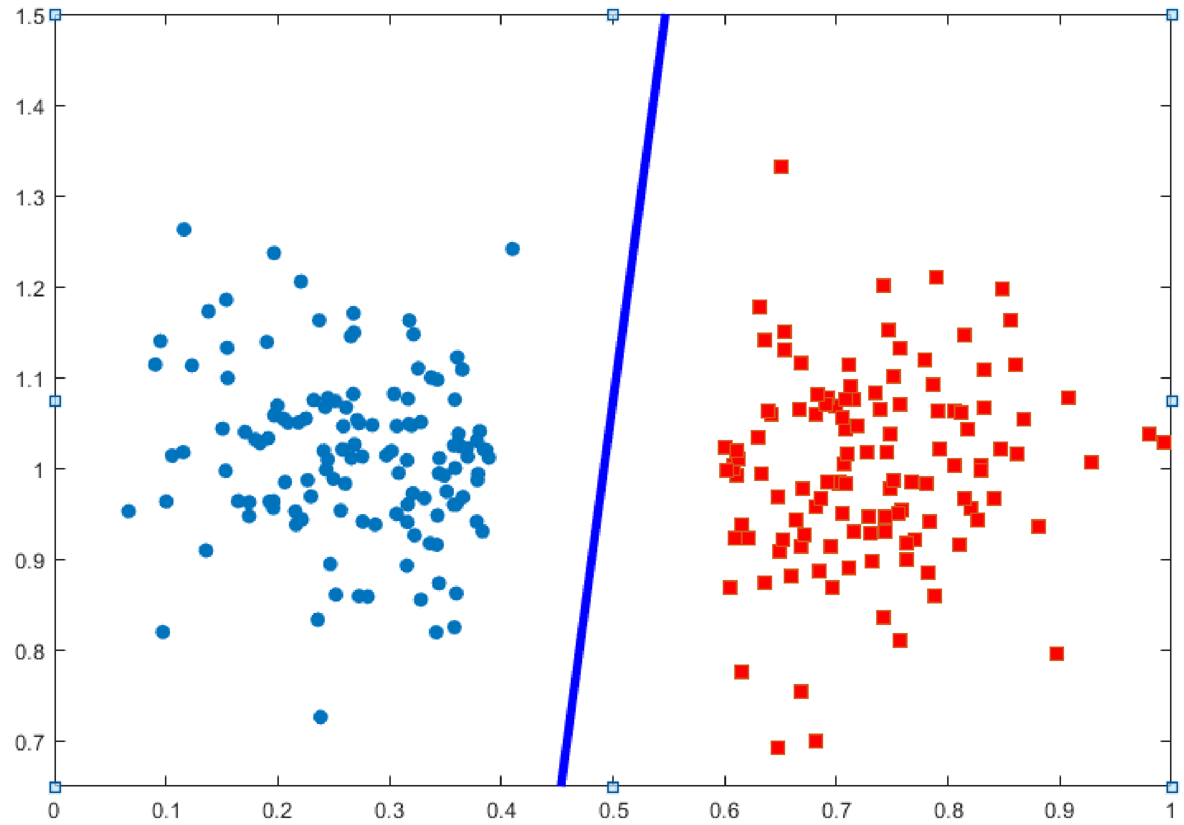


1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1

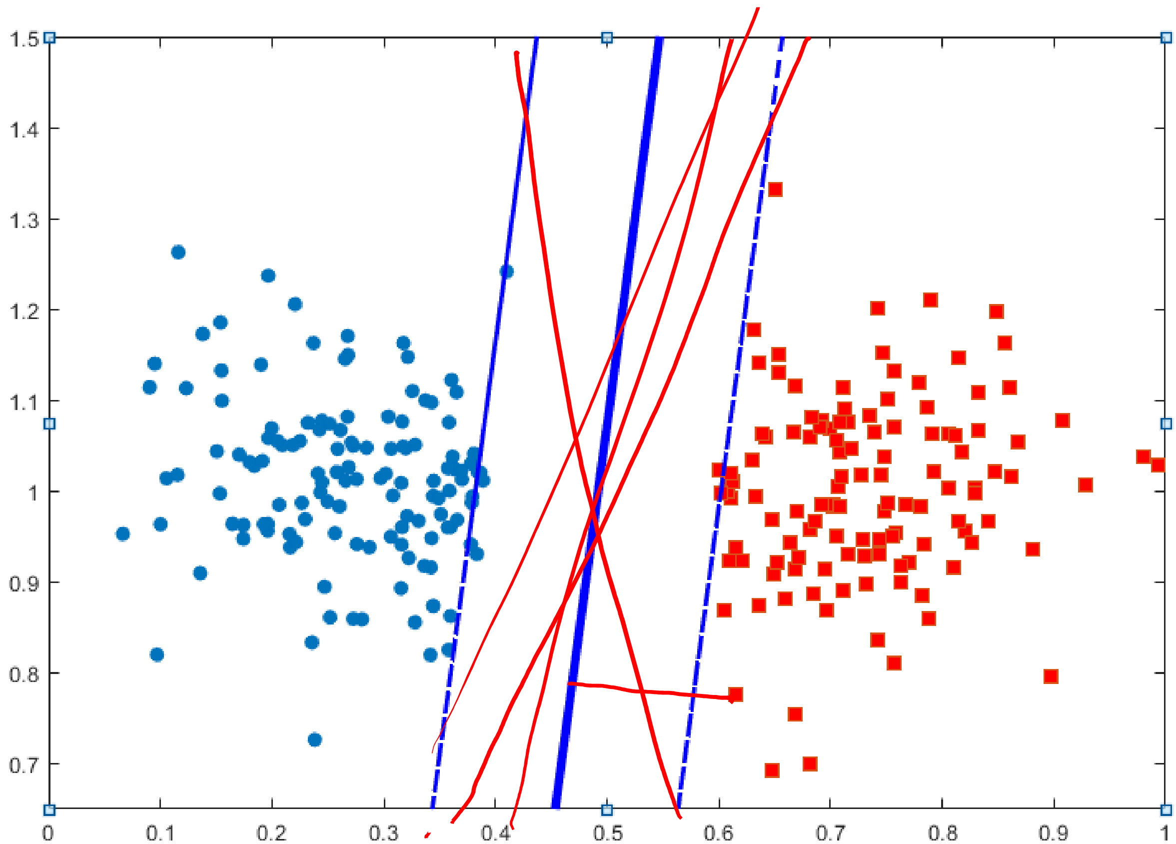

1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1



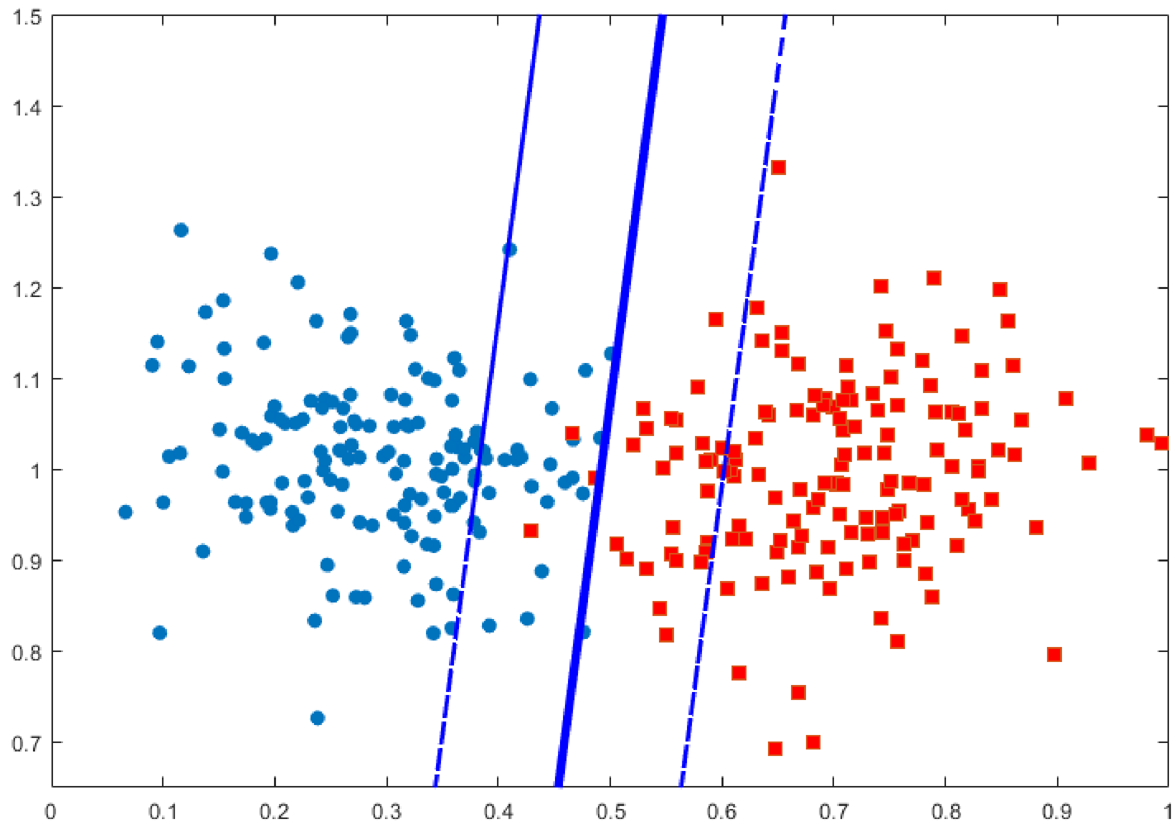
Want to have wider margin ??



1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

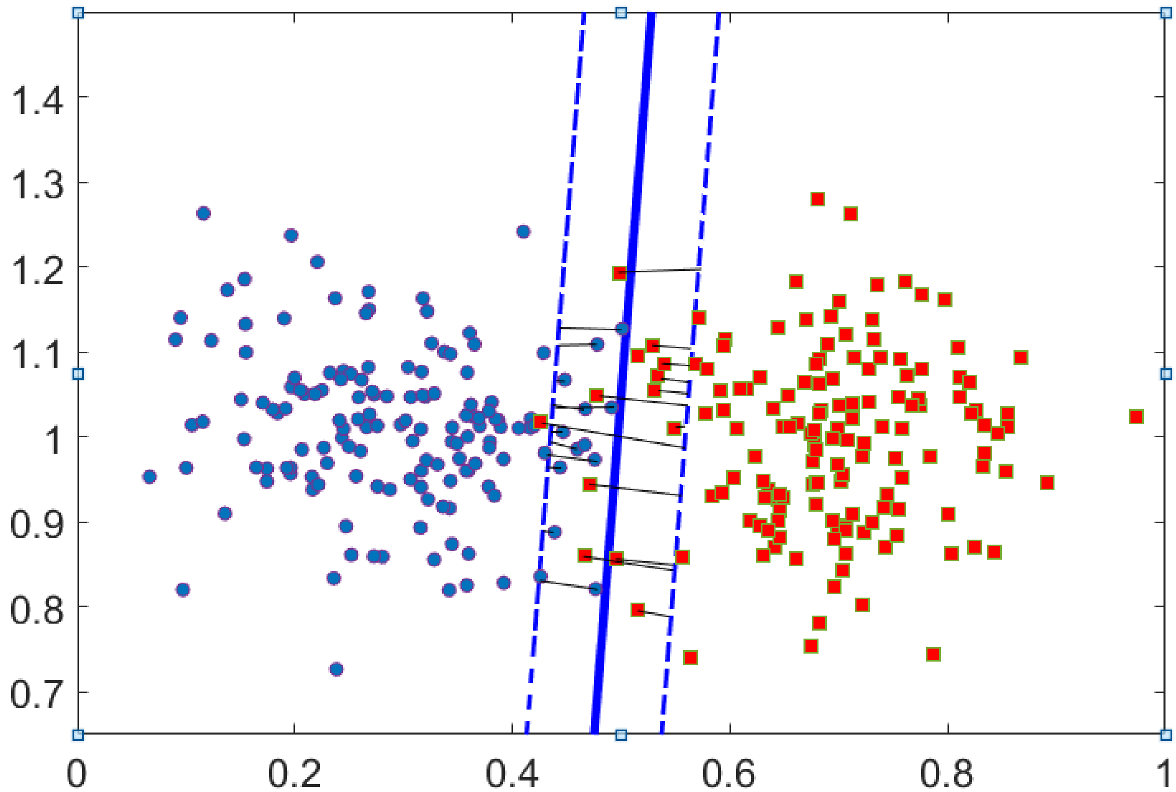


1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

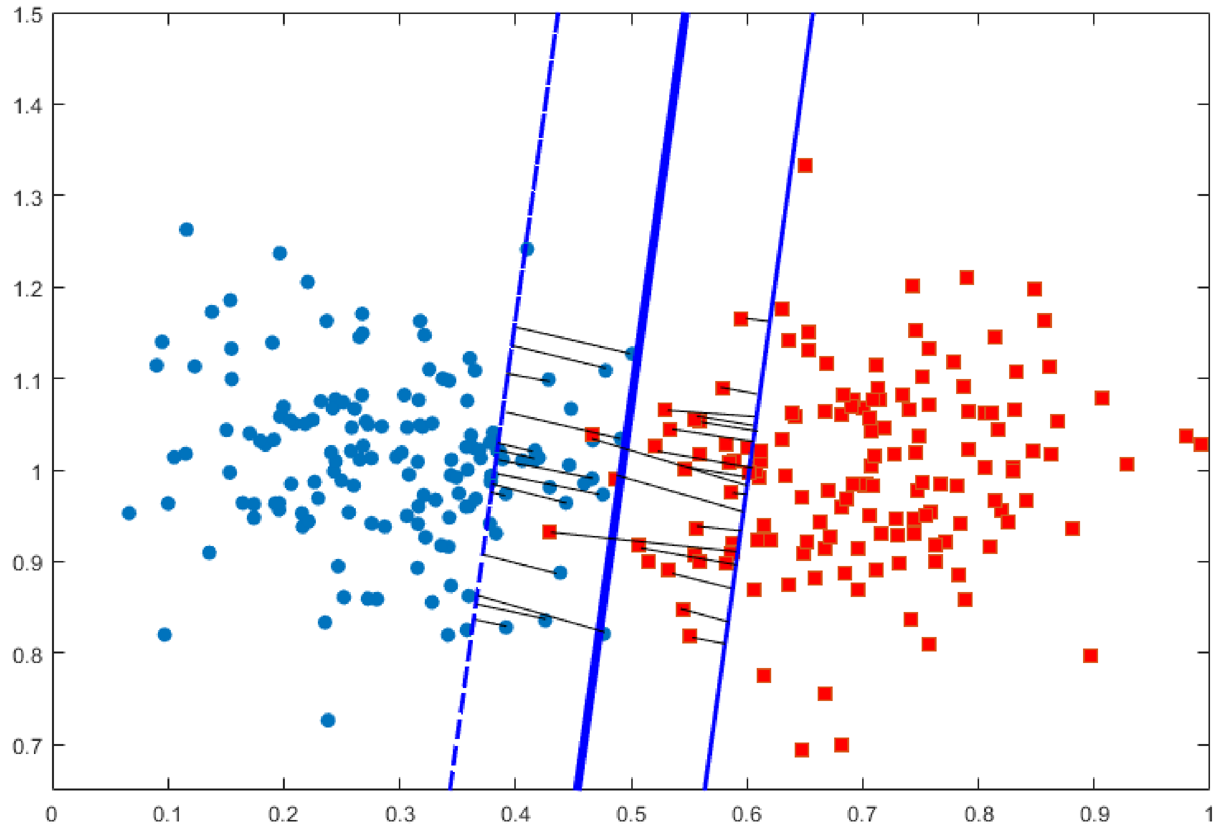


1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

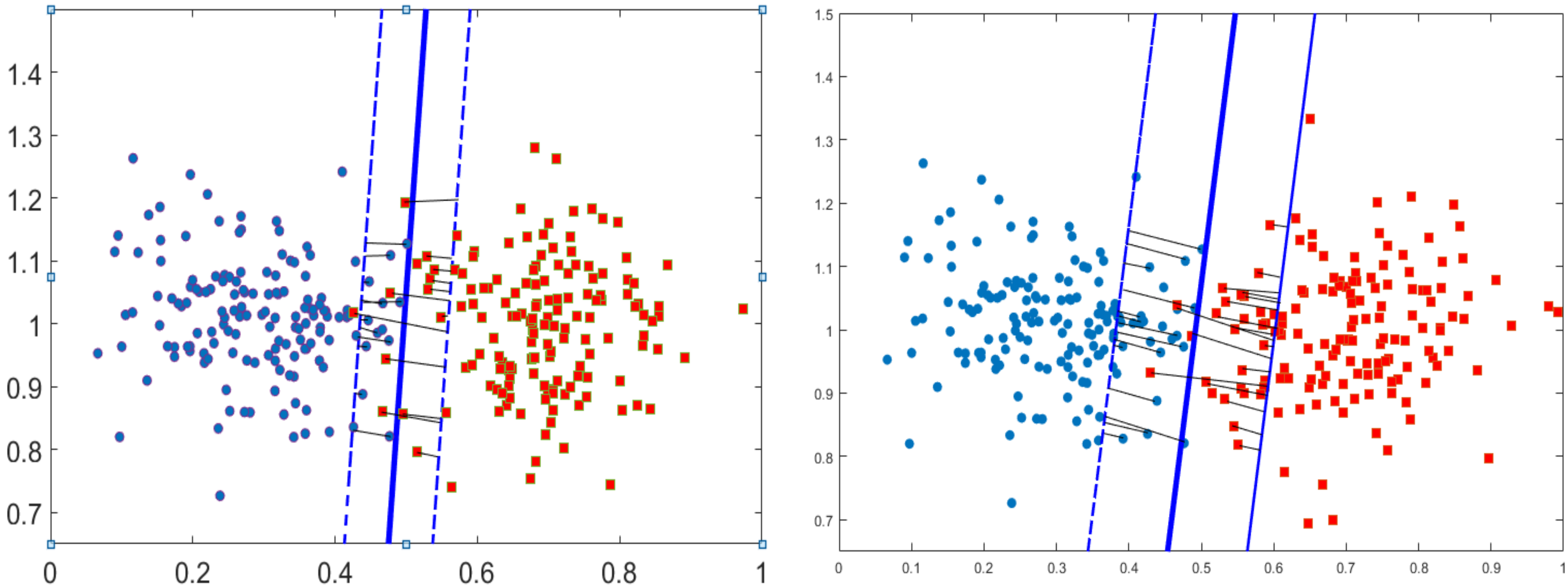
Training Error



Training Error



Increasing width of margin



Observations

Ignore more number of training points

Observations

Ignore more number of training points

Better width of margin

Observations

Ignore more number of training points

Better width of margin

Better separating classification model

Observations

Ignore more number of training points

Better width of margin

Better separating classification model

Larger Training Error

SVM Optimization Problem

Separating hyperplane

$$\beta_1 x_1 + \beta_2 x_2 + \beta_0 = 0$$

$\hat{w}, b$

$[w, b]$

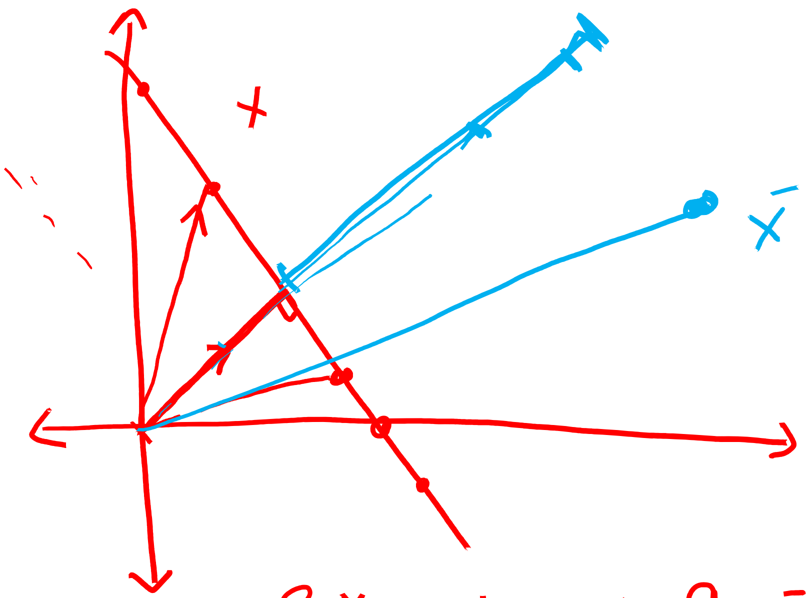
$\hat{w}^T x = c$

$\Rightarrow \left(\frac{w}{\|w\|}\right)^T x = c$

$\Rightarrow w^T x = c\|w\|$

$\Rightarrow w^T x - \underbrace{c\|w\|} = 0$

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



$3x_1 + 4x_2 + 9 = 0$

$1.5x_1 + 2x_2 + 4.5 = 0$

$w^T \hat{x}$

$\hat{w}^T \bar{x} - c$

$\Rightarrow \frac{\hat{w}}{\|w\|} \bar{x} \cancel{-} \frac{c}{\|w\|} \Rightarrow w^T \bar{x} \cancel{-} c$

$$\hat{w}^T x - c$$

$$\hat{w}^T x - \frac{c\,\|w\|}{\|w\|}$$

$$\|w\|\left(\frac{\hat{w}^T x}{\|w\|} - \frac{c\,\|w\|}{\|w\|}\right)$$

$$\frac{1}{\|w\|}\left(w^T x + b\right) = \frac{1}{\|w\|}\left(w^T x + b\right)$$

$\sim$

$$\max_{w,b} \frac{2}{\|w\|} \Big\} \qquad 2$$

Subject to,

$$(w^T x + b) \geq 1 \quad \text{if } y_i = 1$$

$$(w^T x + b) \leq -1 \quad \text{if } y_i = -1$$

$$\underset{w,b}{\overset{\text{Min}}{\cancel{\text{Max}}}} \ \frac{\|w\|^2}{2}$$

Subject to,

$$w^T x + b \geq 1 \quad \text{if } y_i = 1$$

$$w^T x + b \leq -1 \quad \text{if } y_i = -1$$

$$\min_{\substack{w, b \\ w \in \mathbb{R}^n \\ b \in \mathbb{R}}} \quad \frac{1}{2} \cancel{\|w\|^2}\, w^T w$$

Subject to,

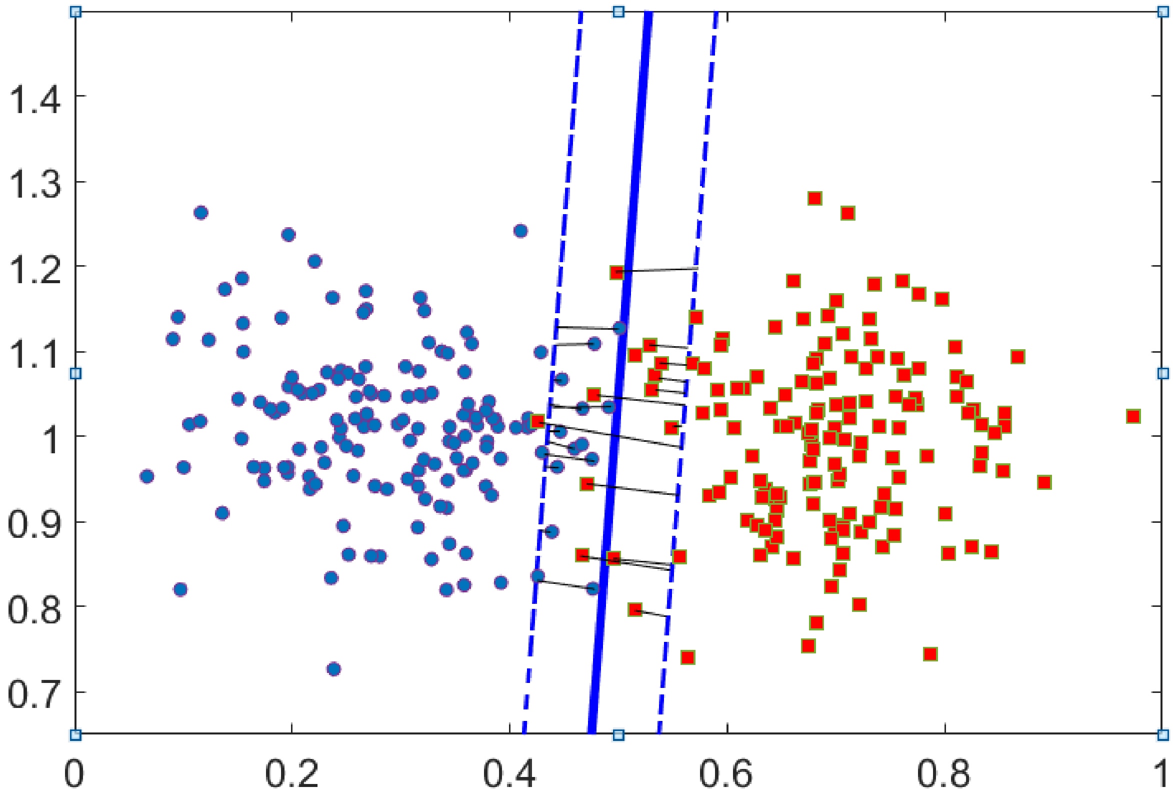
$$y_i(w^T x_i + b) \geq 1; \quad i = 1, 2, \ldots \ell$$

$$\hookrightarrow$$

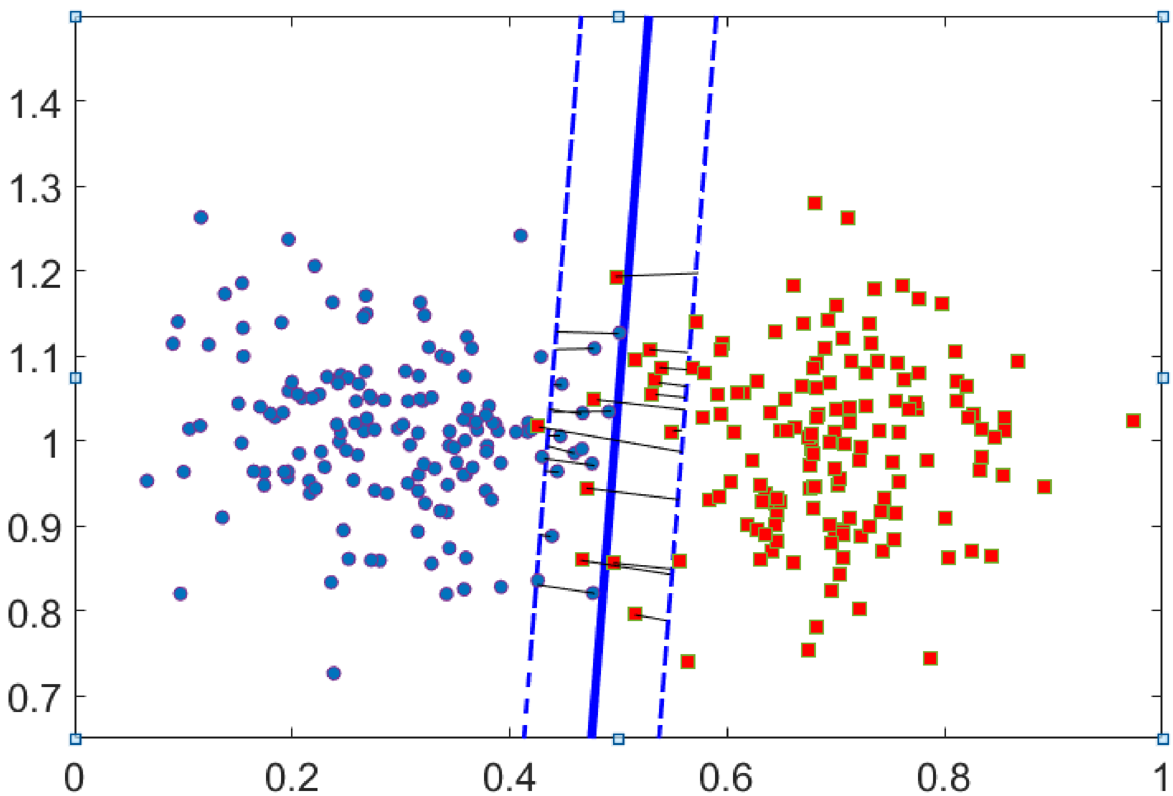
$$\cancel{y} \;(w^T x_i + b) \geq 1 \quad \text{if } y_i = 1$$

$$(w^T x_i + b) \leq -1 \quad \text{if } y_i = -1$$

$$\text{Sign}(w^T x + b)$$



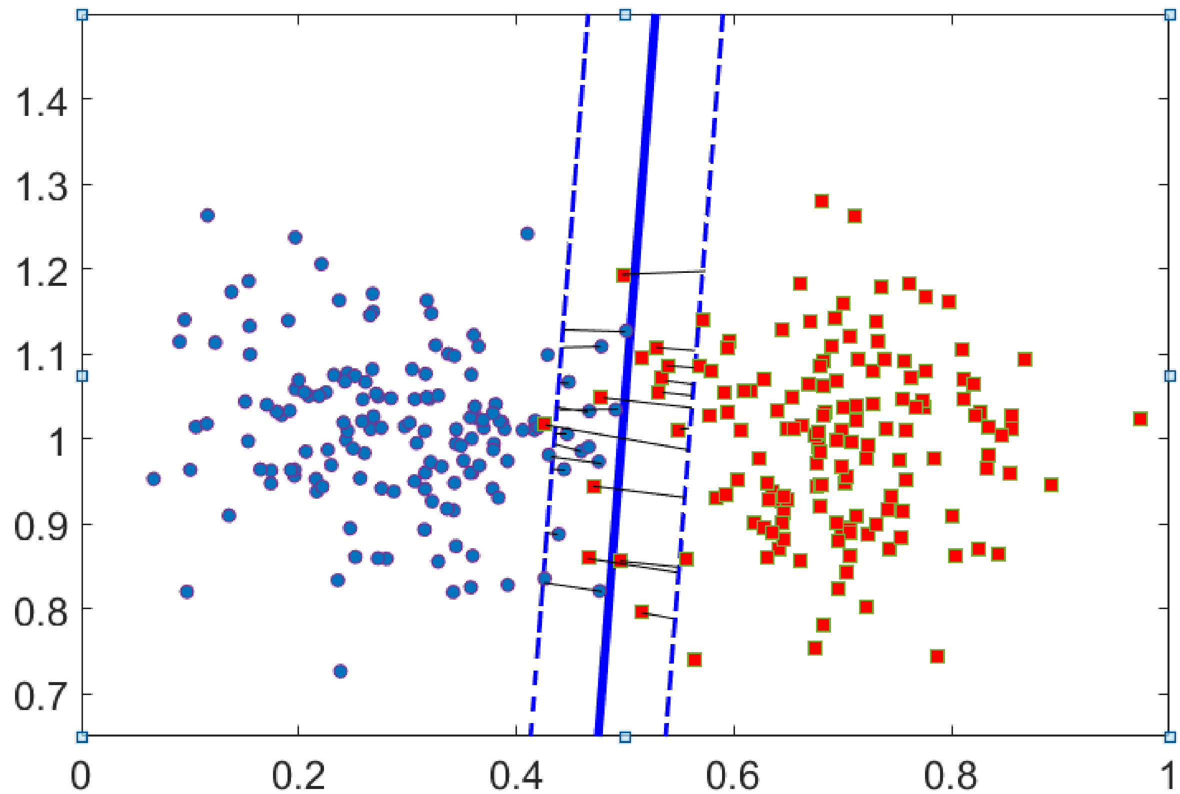
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1



1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1

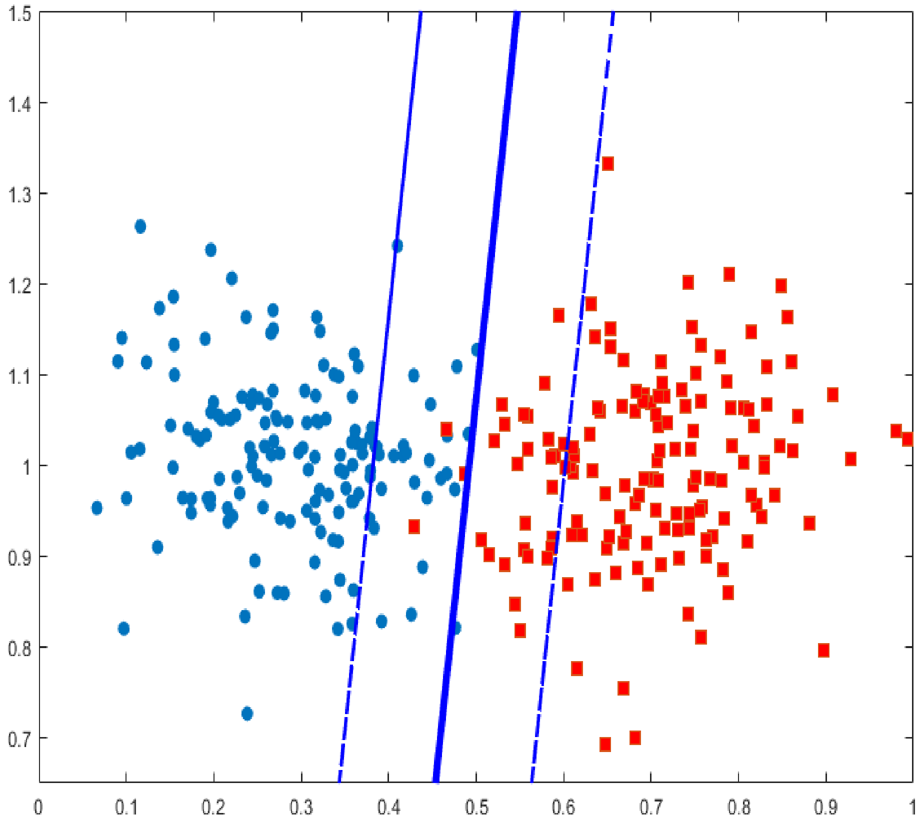
SVM Optimization Problem

$$\text{Min}_{(\beta_{0}, \beta_{1}, \beta_{2})} \quad \underbrace{\frac{1}{2}(\beta_1{}^2 + \beta_2{}^2) \quad +}_{\substack{\text{Model Complexity}\\ \text{Width of margin}}} \quad \underbrace{C\sum_{i=1}^{n}(Max(\, 1 - yi(\, \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_0), 0))}_{\text{Empirical/Training Error}}$$

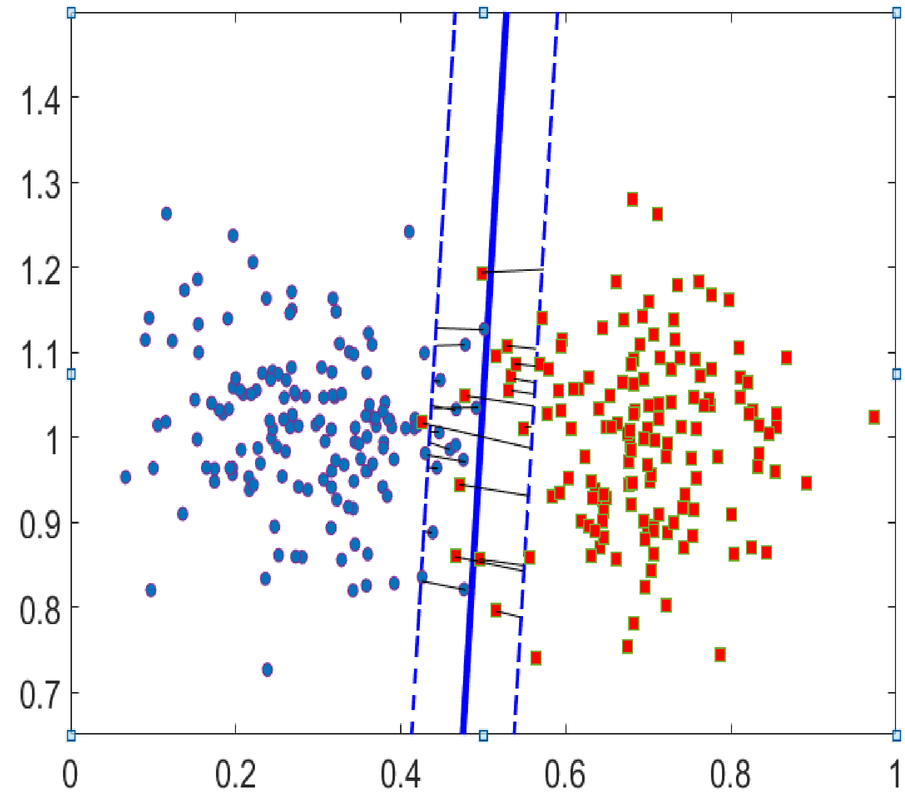

1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1

SVM Optimization Problem

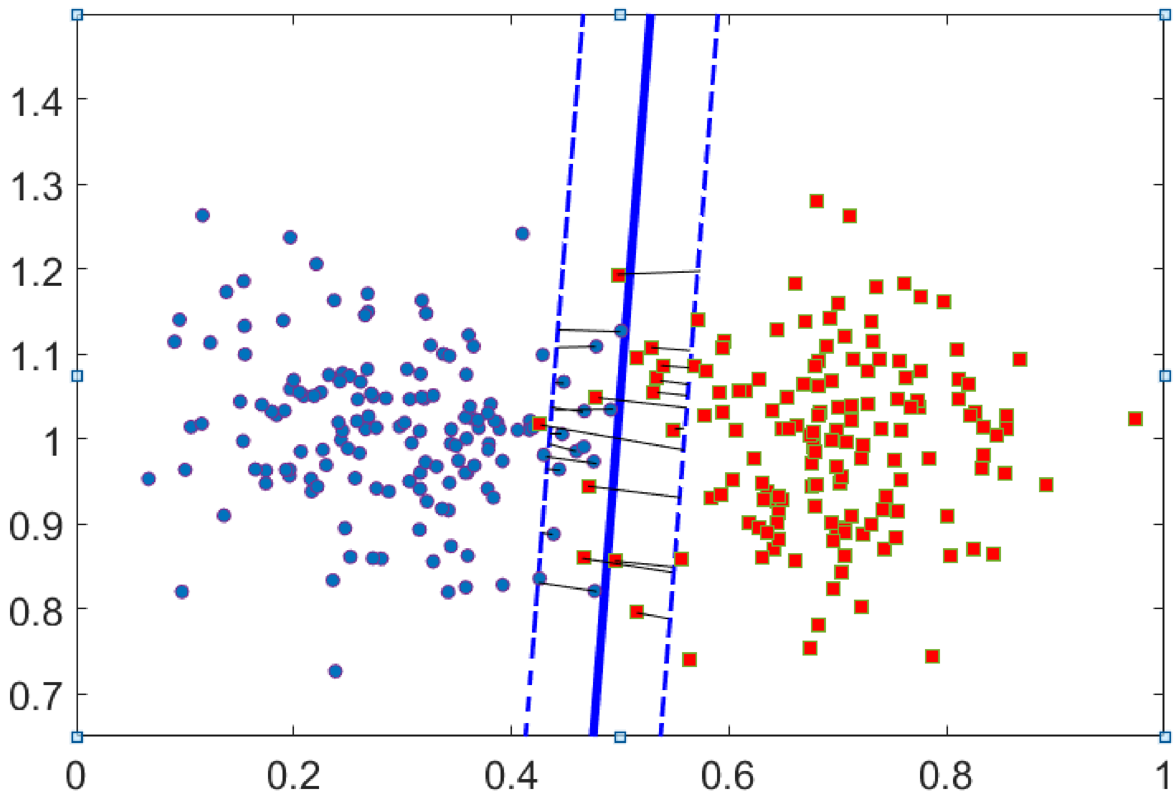
$$\mathrm{Min}_{(\beta_0,\, \beta_1,\, \beta_2)} \quad \underbrace{\frac{1}{2}(\beta_1{}^2 + \beta_2{}^2)}_{\text{Model Complexity}} \quad + \quad \underbrace{C \sum_{i=1}^{n}(Max(\,1 - yi(\,\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_0), 0))}_{\text{Empirical/Training Error}}$$

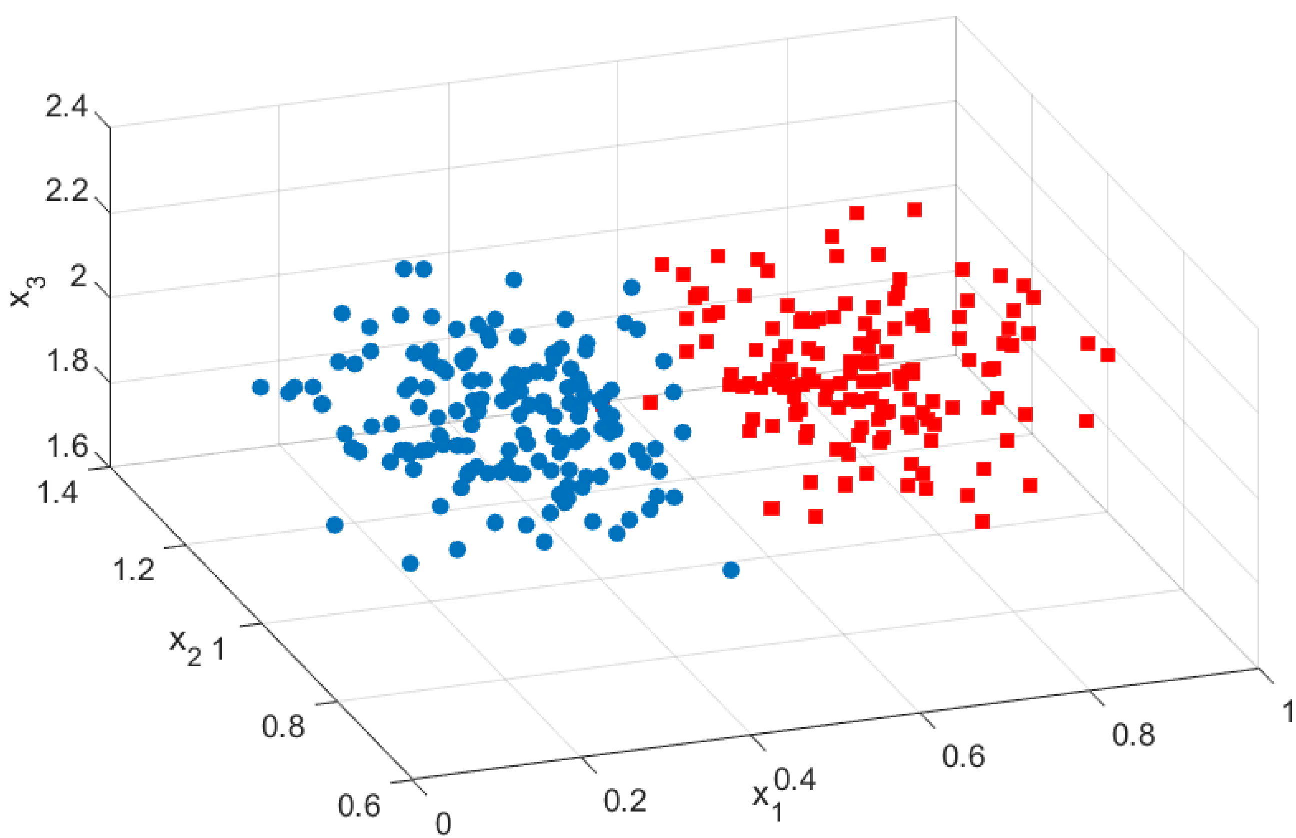

1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

C=0.25

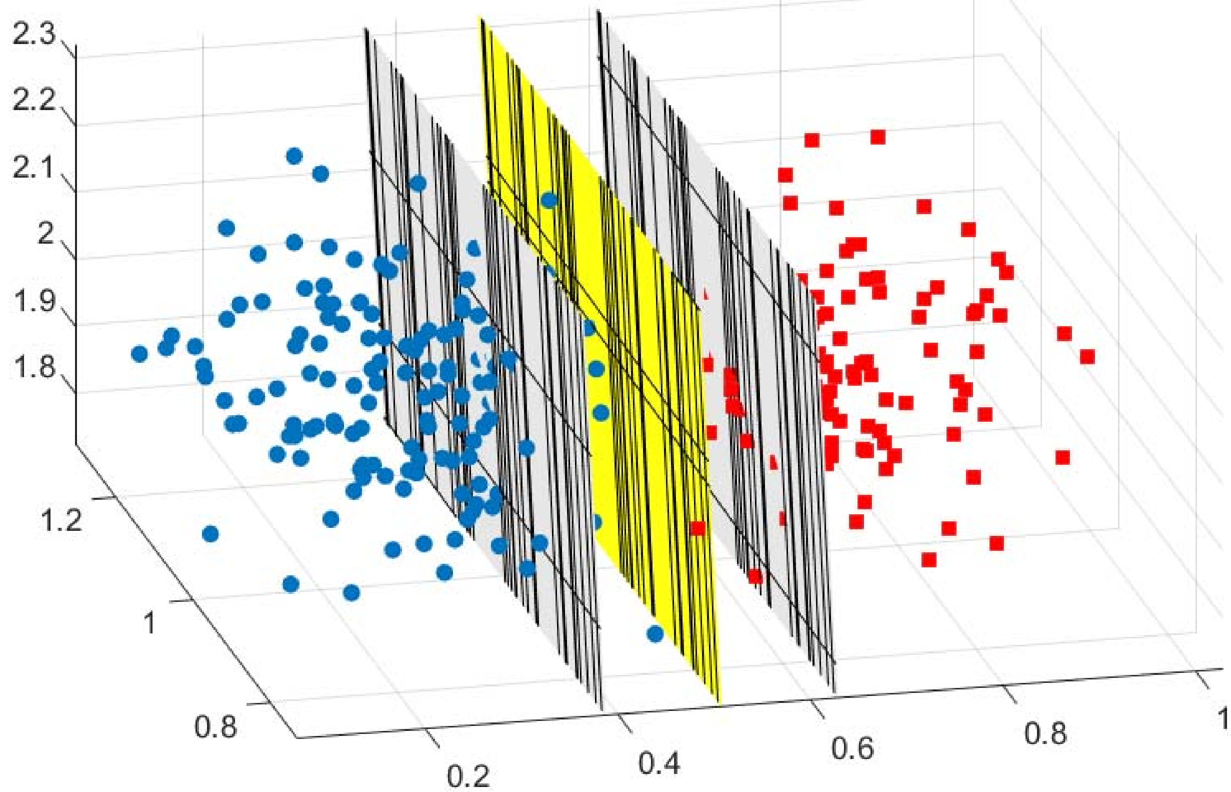

1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1

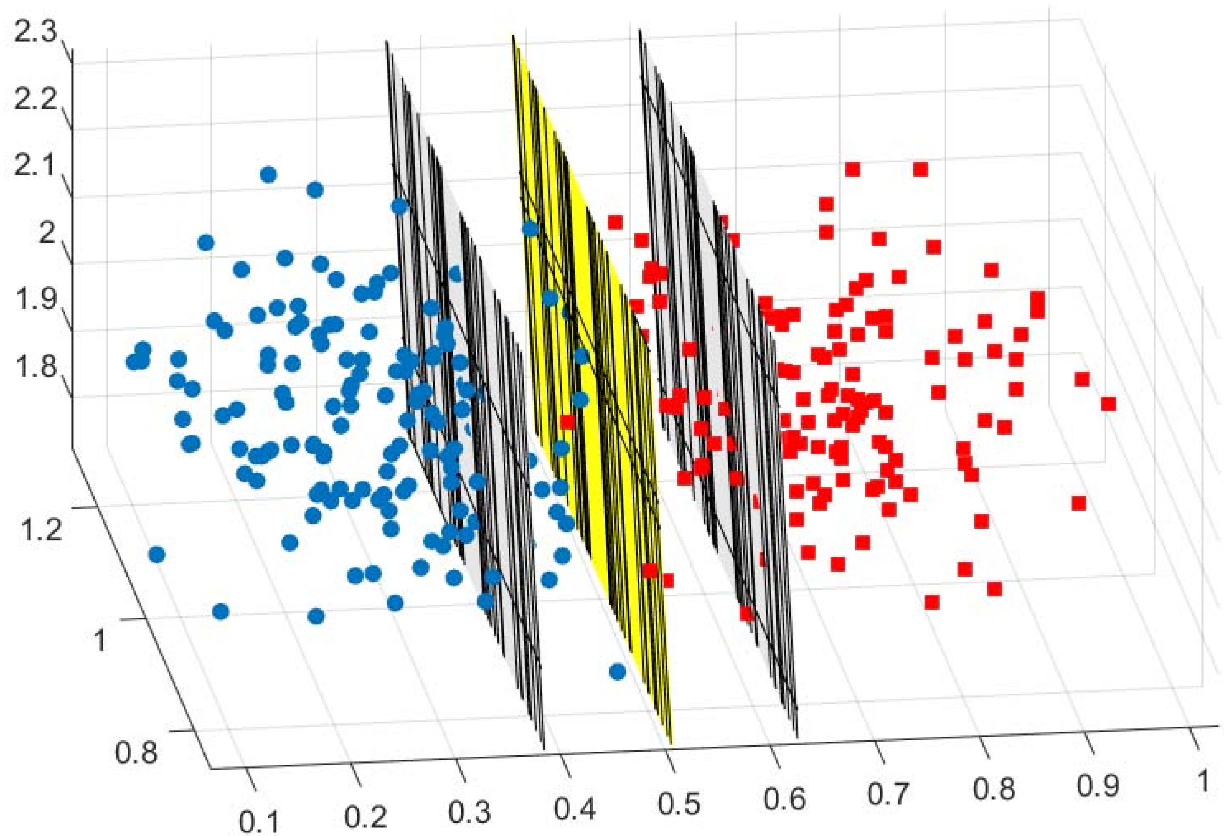
C=4


1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1



2.4
2.2
2
1.8
1.6
x_3
1.4
1.2
x_2 1
0.8
0.6
0
0.2
x_1 0.4
0.6
0.8
1


2.3
2.2
2.1
2
1.9
1.8
1.2
1
0.8
0.2
0.4
0.6
0.8
1


2.3
2.2
2.1
2
1.9
1.8
1.2
1
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

SVM Optimization Problem

Separating hyperplane in 3D

$$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_0 = 0$$

SVM Optimization Problem

$$\text{Min}_{(\beta_0, \beta_1 \beta_2, \beta_3)} \quad \underbrace{\frac{1}{2}(\beta_1{}^2 + \beta_2{}^{\mathbf{2}} + \beta_3{}^2)}_{\text{Model Complexity}} \quad + \quad C \underbrace{\sum_{i=1}^{n}(Max(\,1 - yi(\,\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_0), 0))}_{\text{Empirical/Training Error}}$$

SVM Optimization Problem

$$\text{Min}_{(\beta_{0,}\ \beta_1\ \beta_2,\ \beta_3)} \quad \underbrace{\frac{1}{2}(\beta_1{}^2 + \beta_2{}^2 + \beta_3{}^2)}_{\text{Model Complexity}} \; + \; C \sum_{i=1}^{n} \underbrace{(Max(\,1 - yi(\,\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_0), 0))}_{\text{Empirical/Training Error}}$$

SVM Optimization Problem

Separating hyperplane in general

$$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots + \beta_n x_n + \beta_0 = 0$$

SVM Optimization Problem

$$\text{Min}_{(\beta_0,\ \beta_1,\ \beta_2,\ldots,\beta_n)} \quad \underbrace{\frac{1}{2}\left(\beta_1^{\,2} + \beta_2^{\,2} + \beta_3^{\,2} + \ldots + \beta_n^{\,2}\right)}_{\text{Model Complexity}} \; + \underbrace{C \sum_{i=1}^{n} \left(Max\left(1 - yi\left(\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_n x_{ni} + \beta_0\right), 0\right)\right)}_{\text{Empirical/Training Error}}$$

SVM Optimization Problem

$$\text{Min}_{(w \in R^n, b \in R)} \quad \underbrace{\frac{1}{2}(w^T w)}_{\text{Model Complexity}} \quad + \quad \underbrace{C \sum_{i=1}^{n}(Max(1 - yi(w^T x_i + b), 0))}_{\text{Empirical/Training Error}}$$

$$w = \begin{bmatrix} \beta n \\ \beta_{n-1} \\ .. \\ \beta 2 \\ \beta_1 \end{bmatrix} \qquad b = \beta_0$$

SVM Primal Optimization Problem

$$\text{Min}_{(w \in R^n, b \in R)} \quad \frac{1}{2}(w^T w) \quad + \quad C \sum_{i=1}^{n} \xi_i$$

Subject to,

$$y_i(w^T x_i + b) \geq 1 - \xi_i,$$
$$\xi_i > 0, i = 1,2, \dots, l.$$

SVM Dual Optimization Problem

$$\mathrm{M}ax_{(\alpha \in R^n)} \quad -\frac{1}{2}\left(\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j \, y_i y_j {x_i}^T x_j\right) \quad + \sum_{i=1}^{n} \alpha_{i\,i}$$

Subject to,

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

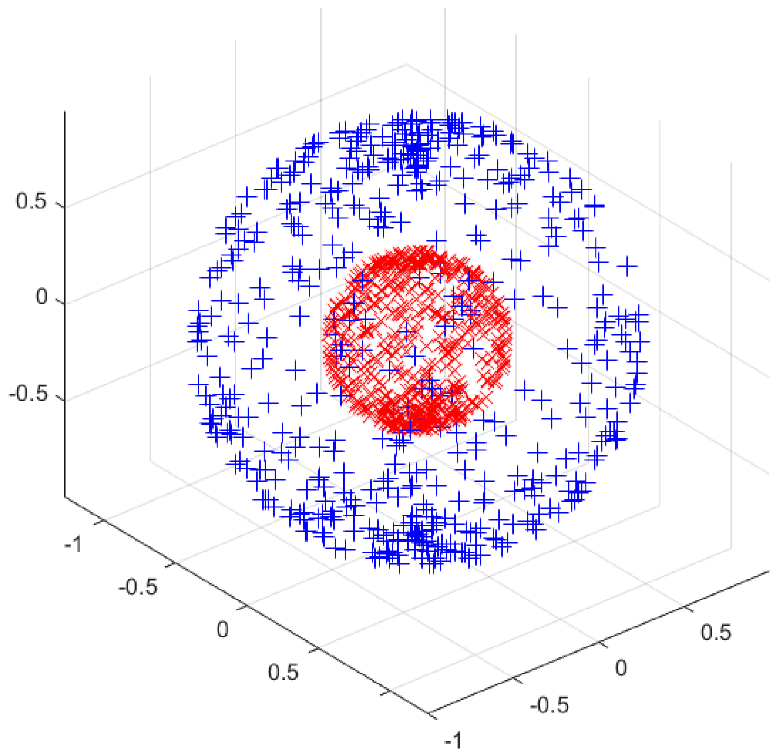
$$0 \le \alpha_i \le C,$$ i =1,2,…C

SVM Optimization Problem

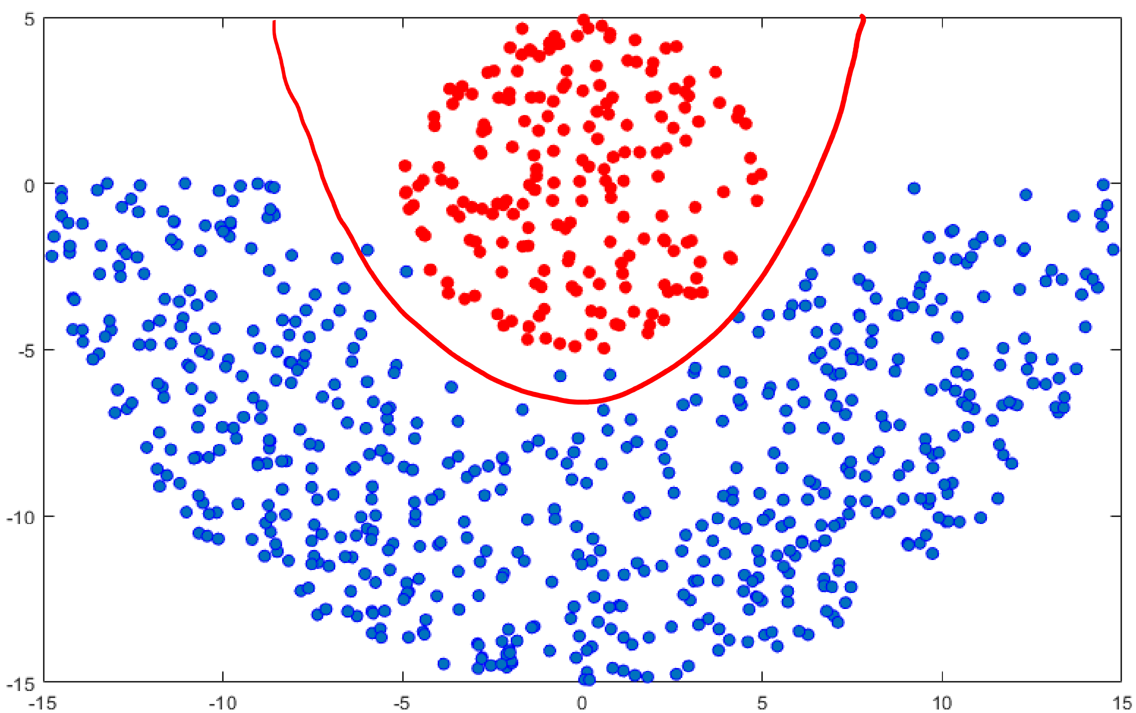
$$\text{Min}_{(w \in R^n, b \in R)} \quad \underbrace{\frac{1}{2}(w^T w)}_{\text{Model Complexity}} \quad + \quad \underbrace{C \sum_{i=1}^{n} (Max(\,1 - yi(\,w^T x_i + b), 0))}_{\text{Empirical/Training Error}}$$

$$w = \begin{bmatrix} \beta n \\ \beta_{n\,-\,1} \\ .. \\ \beta 2 \\ \beta_1 \end{bmatrix} \qquad b = \beta_0$$

Non-linear separable data

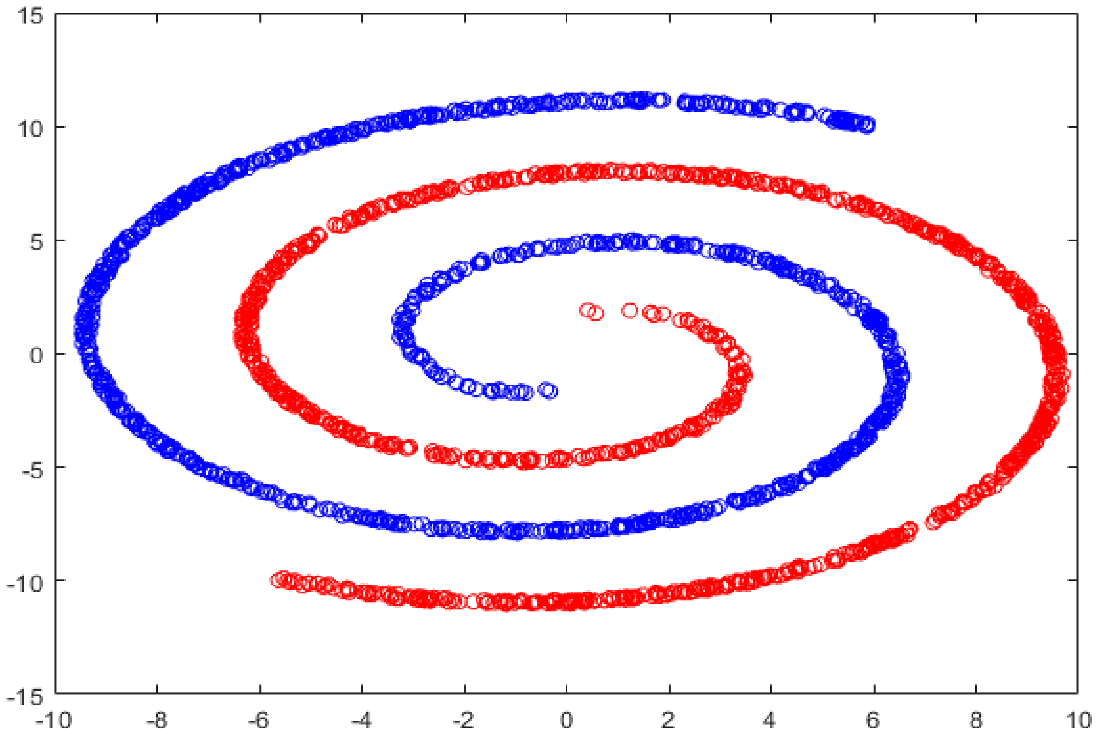


Non-linear separable data



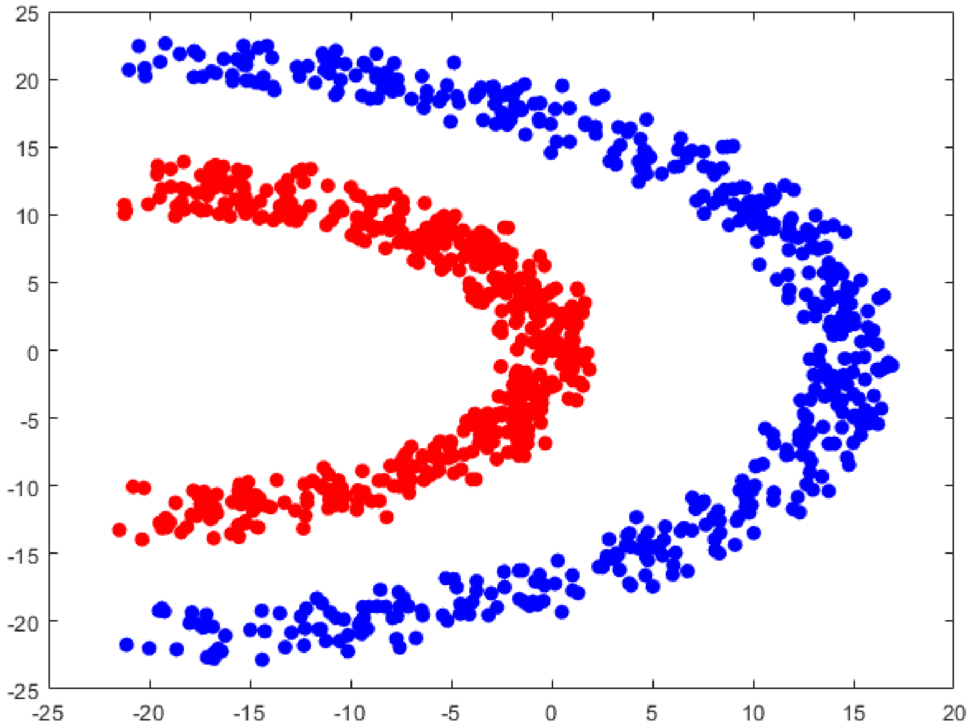
Crescent full moon dataset

Non-linear separable data



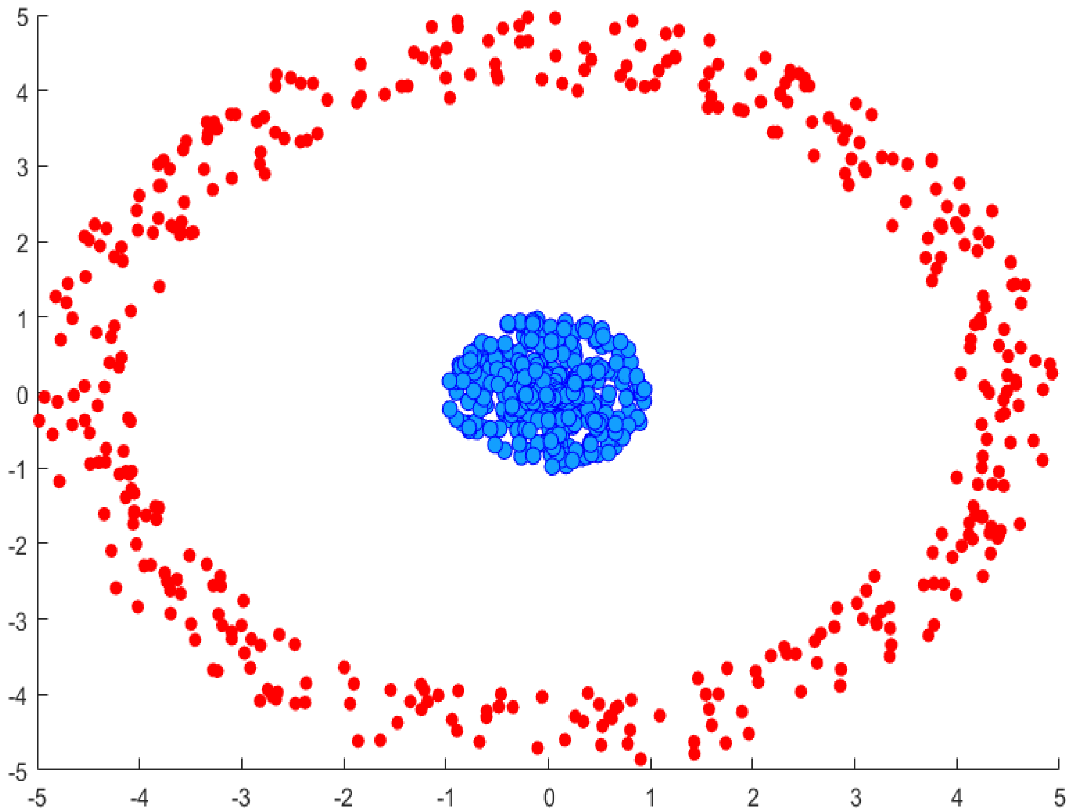
Spiral Dataset

Non-linear separable data



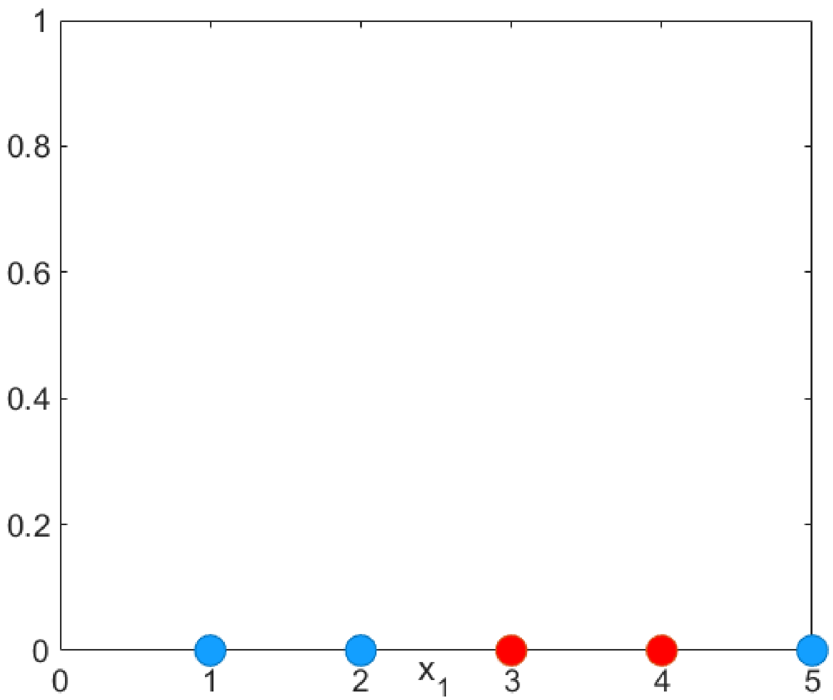
Half kernel Dataset

Non-linear separable data

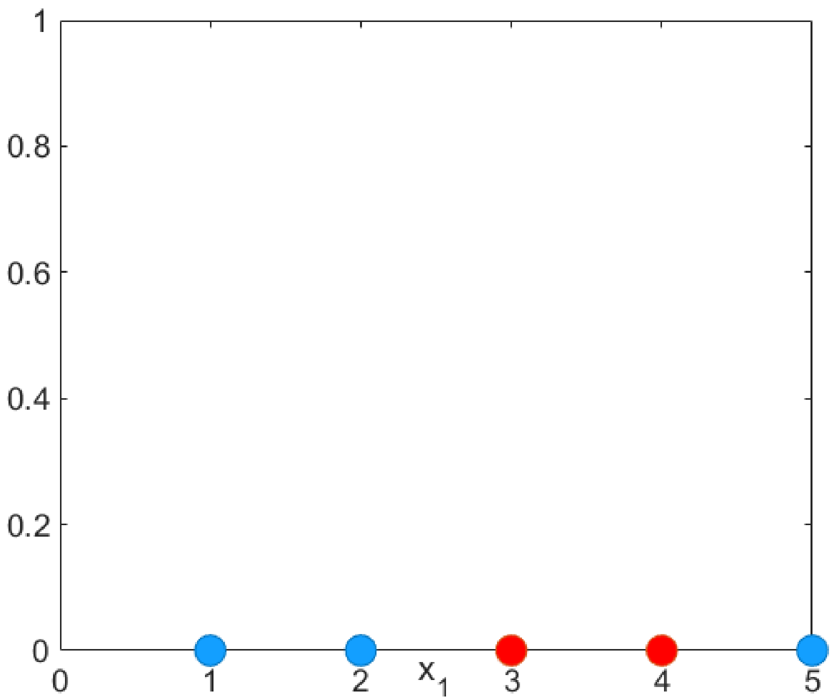


Co-centric Circle Dataset

Kernel Trick



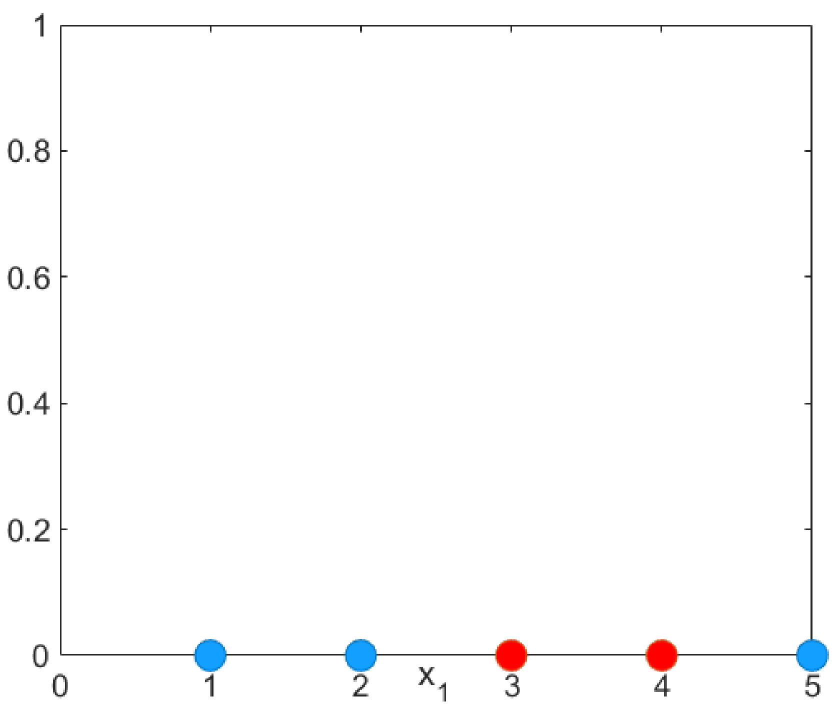
Kernel Trick



Φ(x)

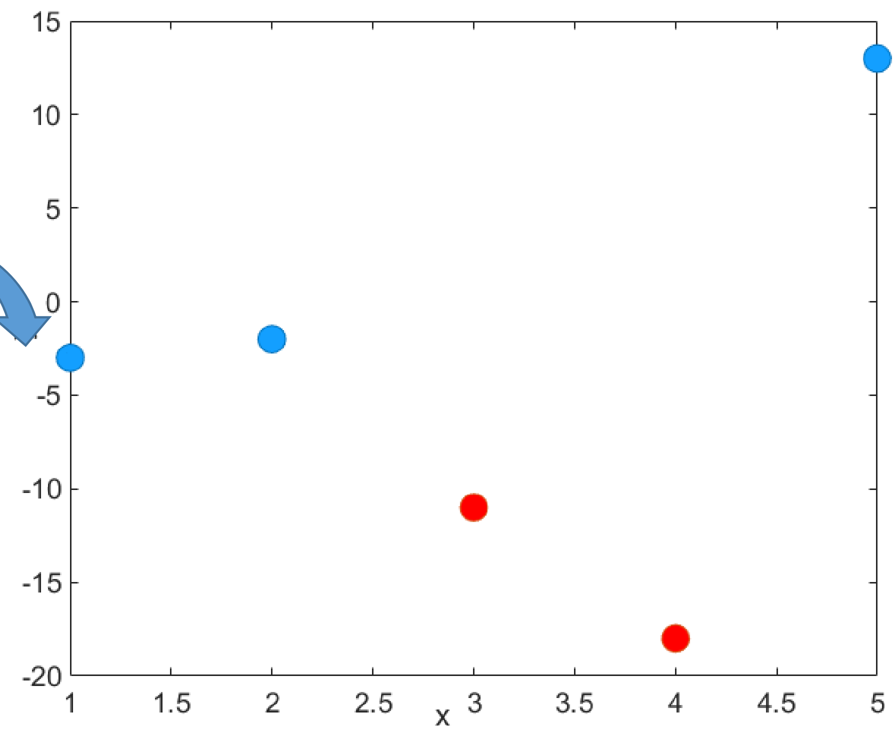
$$= \begin{bmatrix} x \\ x^2 - 2x - 2 \end{bmatrix}$$

Kernel Trick

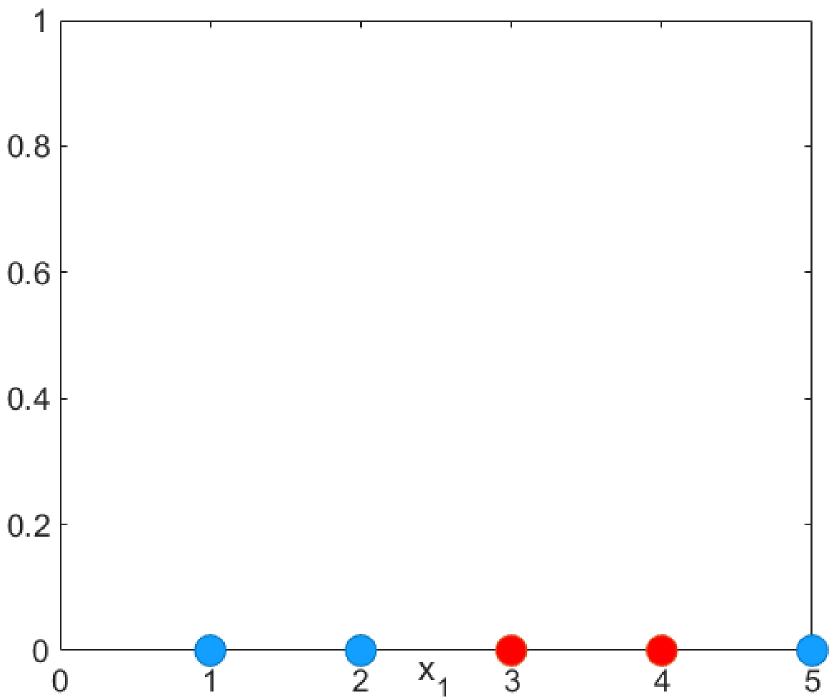


Φ(x)

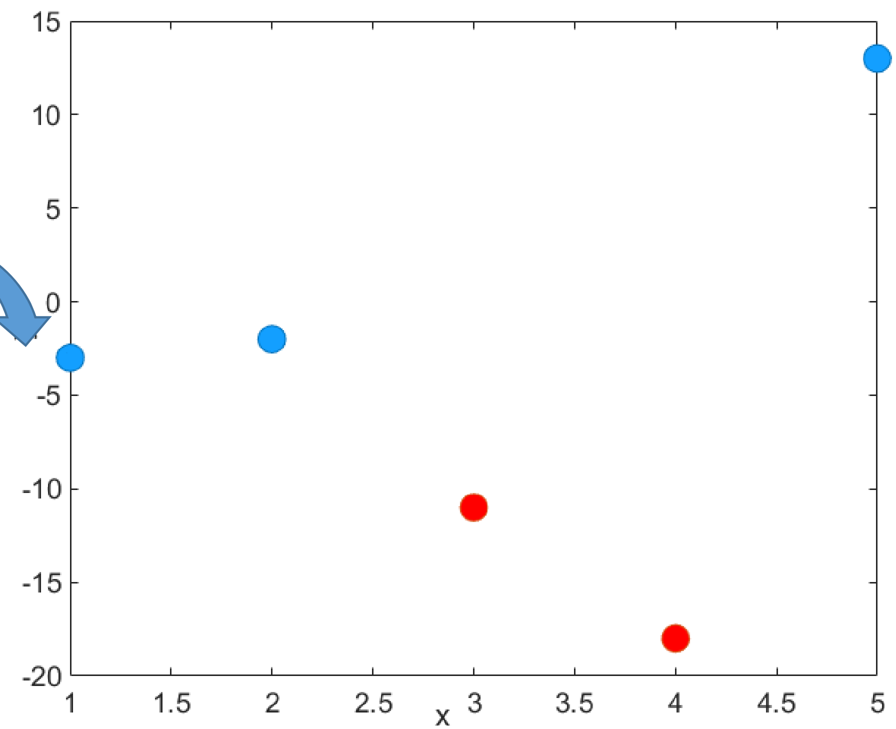
$$= \begin{bmatrix} x \\ x^2 - 2x - 2 \end{bmatrix}$$



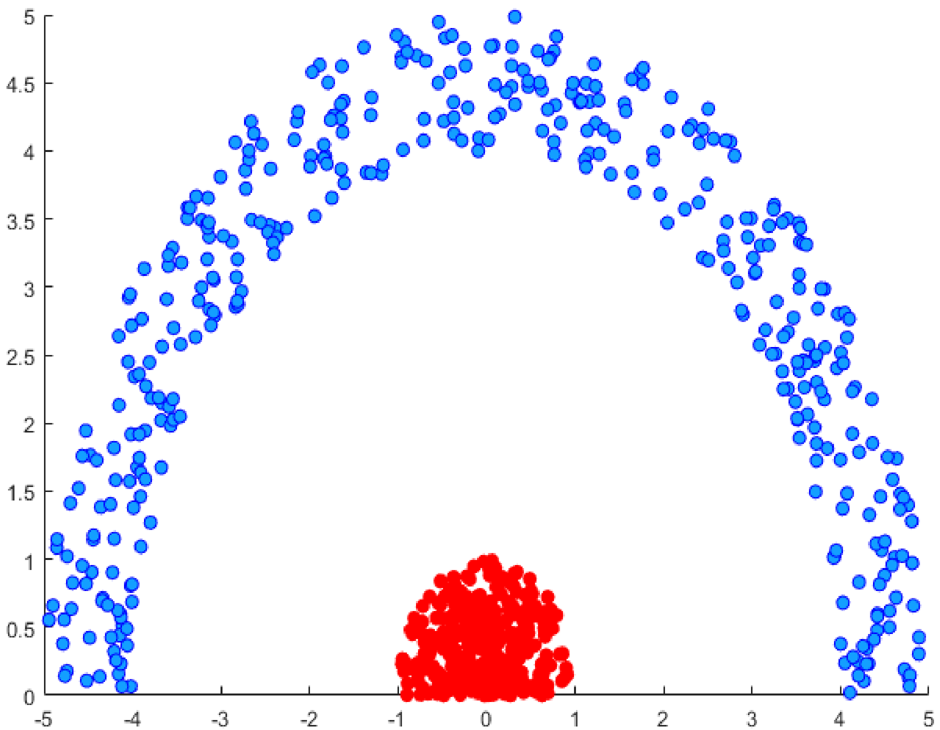
Kernel Trick



$$\Phi(x) = \begin{bmatrix} x \\ x^2 - 2x - 2 \end{bmatrix}$$

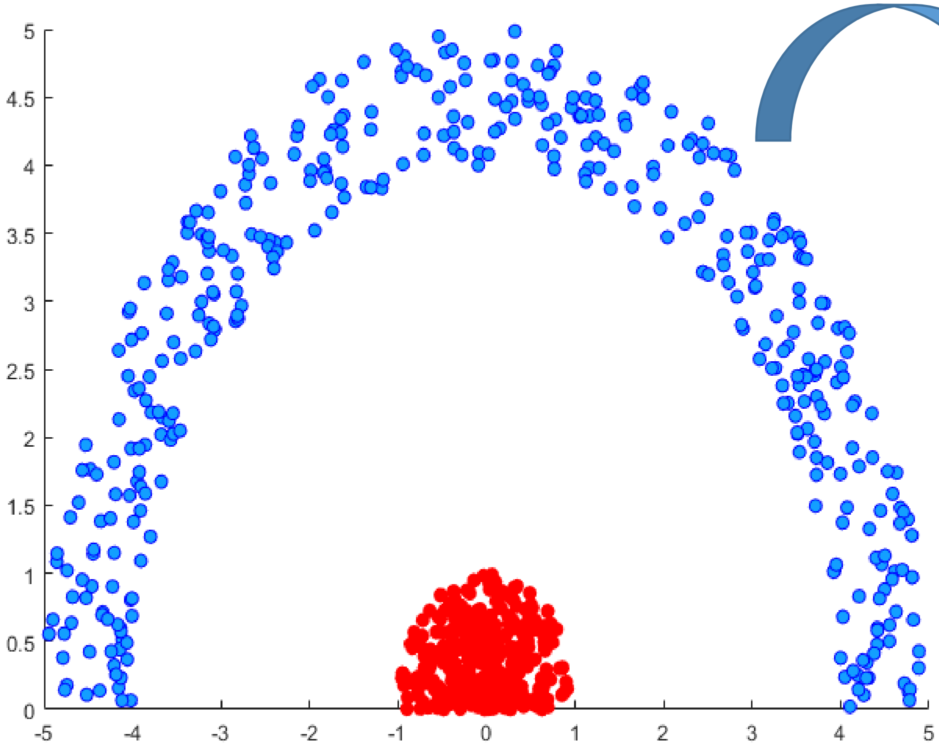


Kernel Trick



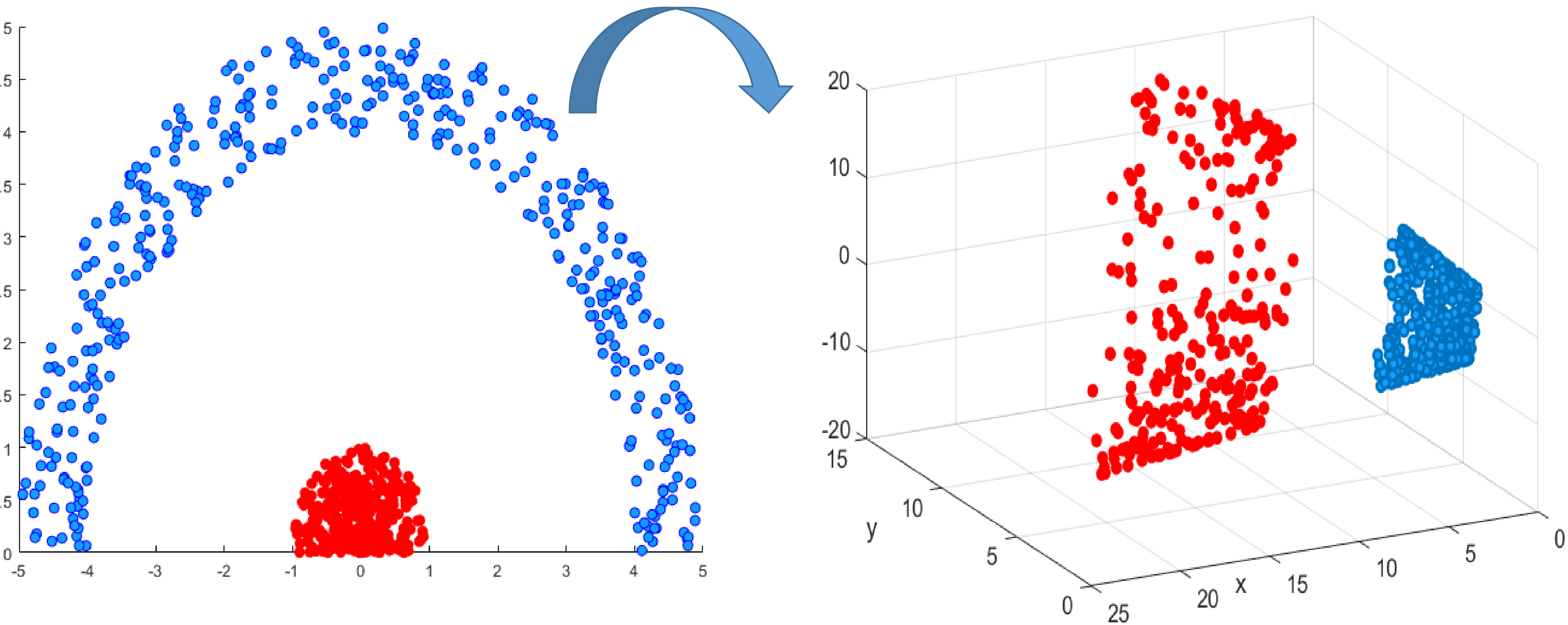
Kernel Trick

$$\Phi(x) = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} {x_1}^2 \\ {x_2}^2 \\ \sqrt{2}x_2x_1 \end{bmatrix}$$



Kernel Trick

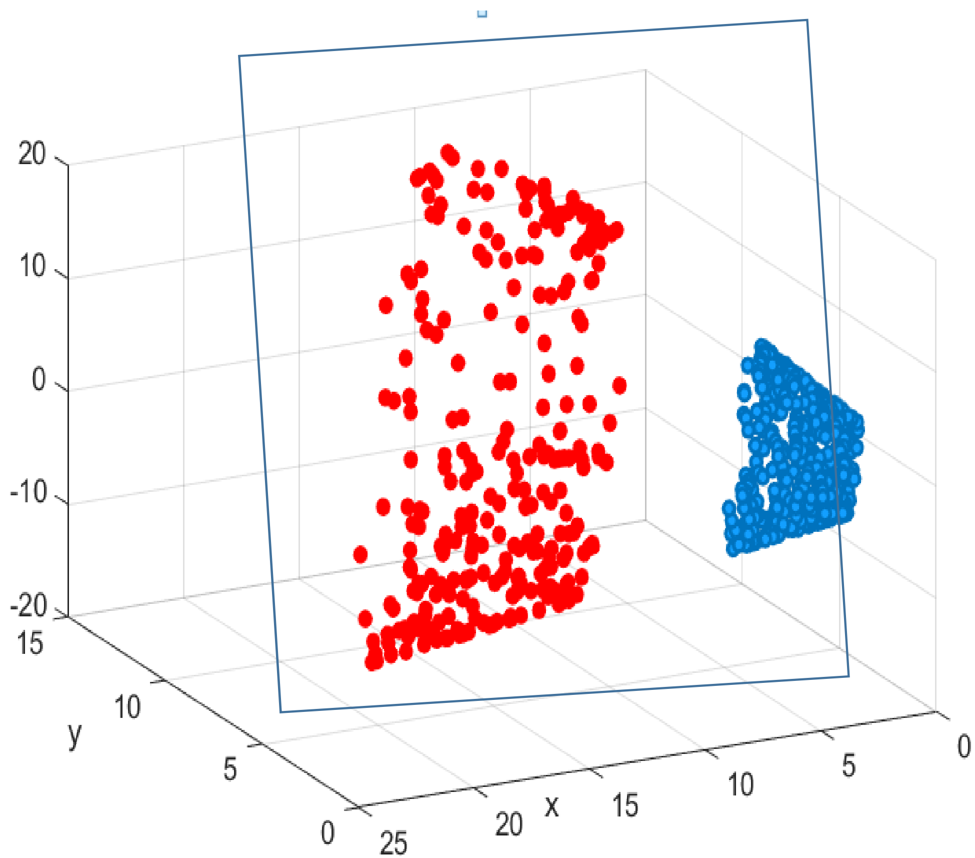
$$\Phi(x) = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} {x_1}^2 \\ {x_2}^2 \\ \sqrt{2}x_2x_1 \end{bmatrix}$$

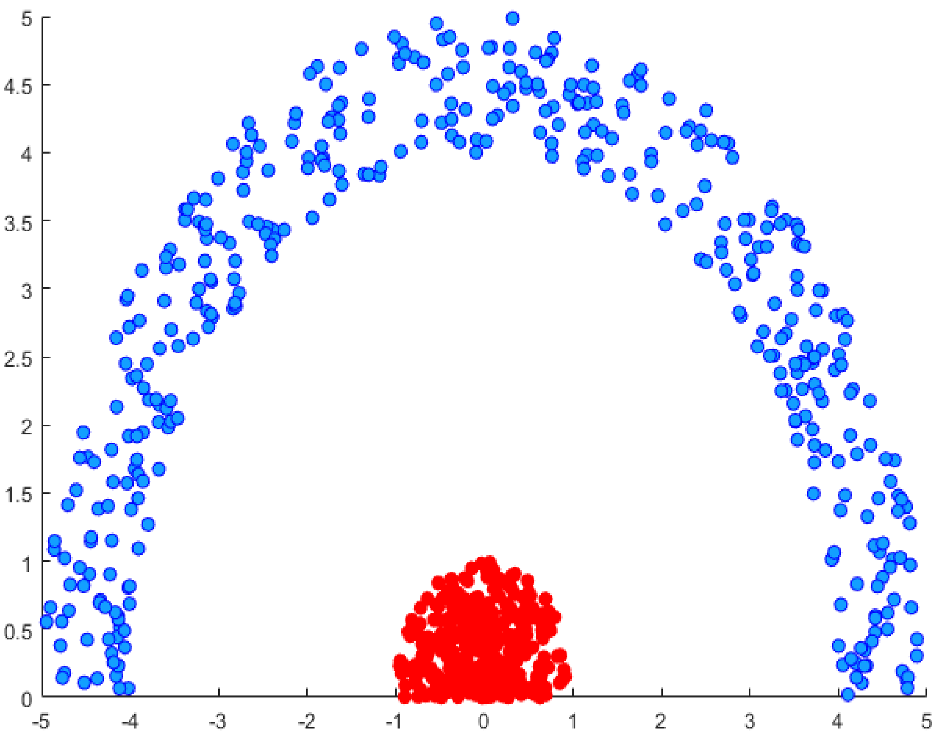


Observations??

Kernel Trick

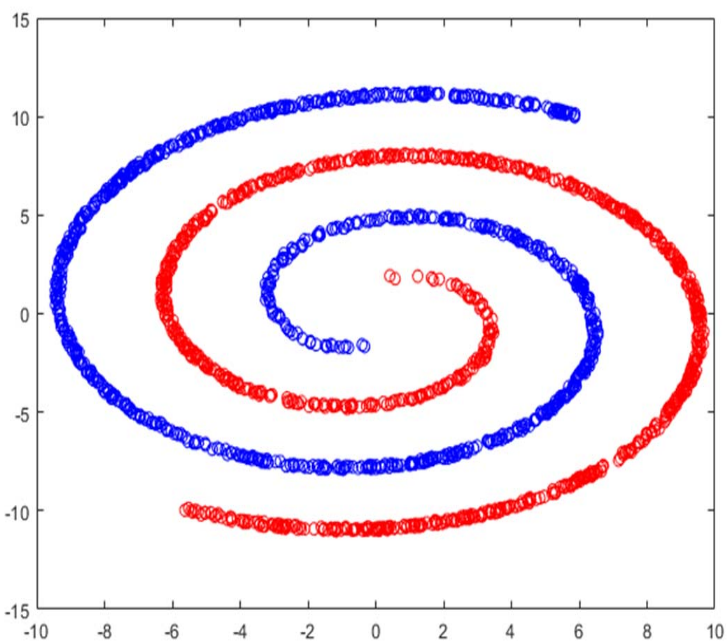
$$\Phi(x) = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} {x_1}^2 \\ {x_2}^2 \\ \sqrt{2}x_2x_1 \end{bmatrix}$$





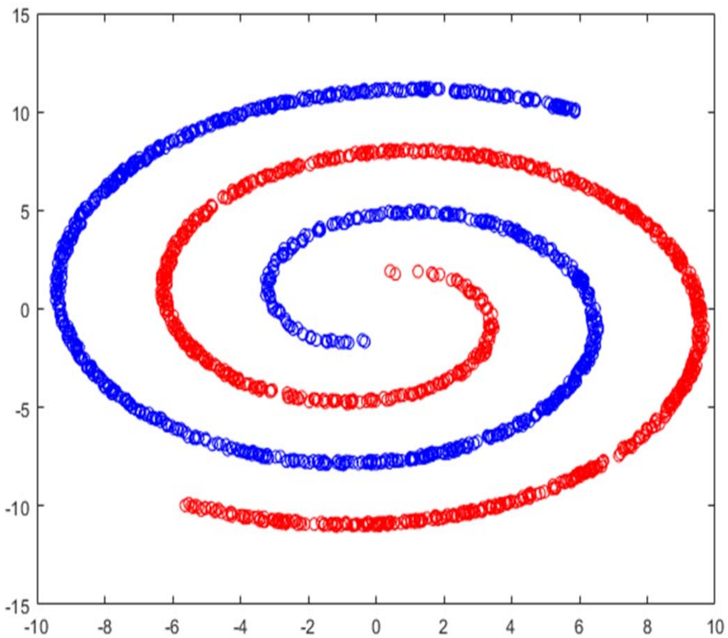
Same mapping for spiral dataset

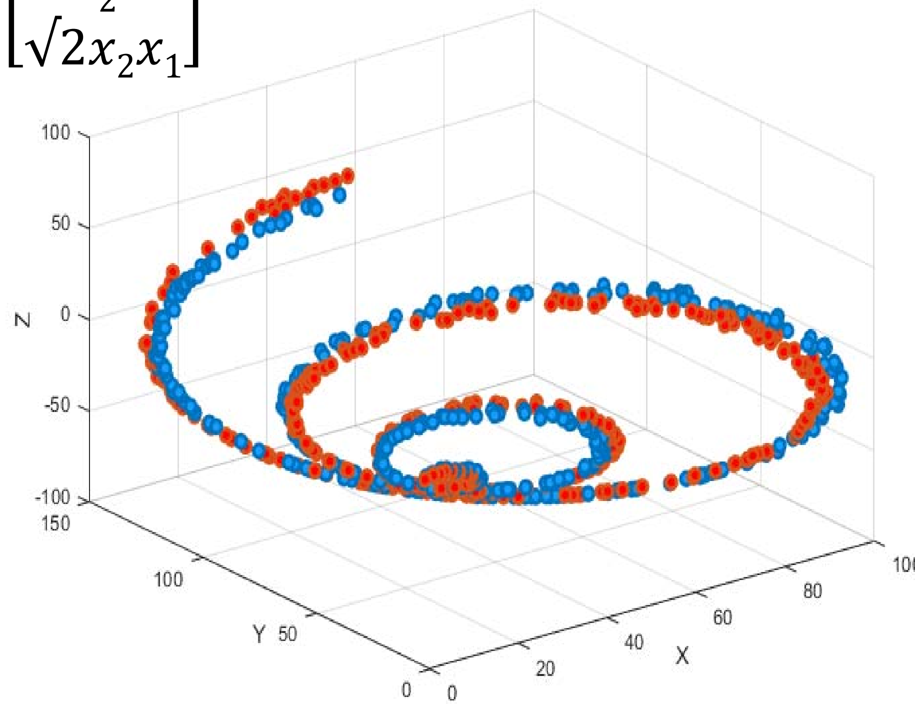
$$\Phi(x) = \Phi\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}{x_1}^2\\{x_2}^2\\\sqrt{2}x_2x_1\end{bmatrix}$$



Same mapping for spiral dataset

$$\Phi(x) = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} {x_1}^2 \\ {x_2}^2 \\ \sqrt{2}x_2x_1 \end{bmatrix}$$



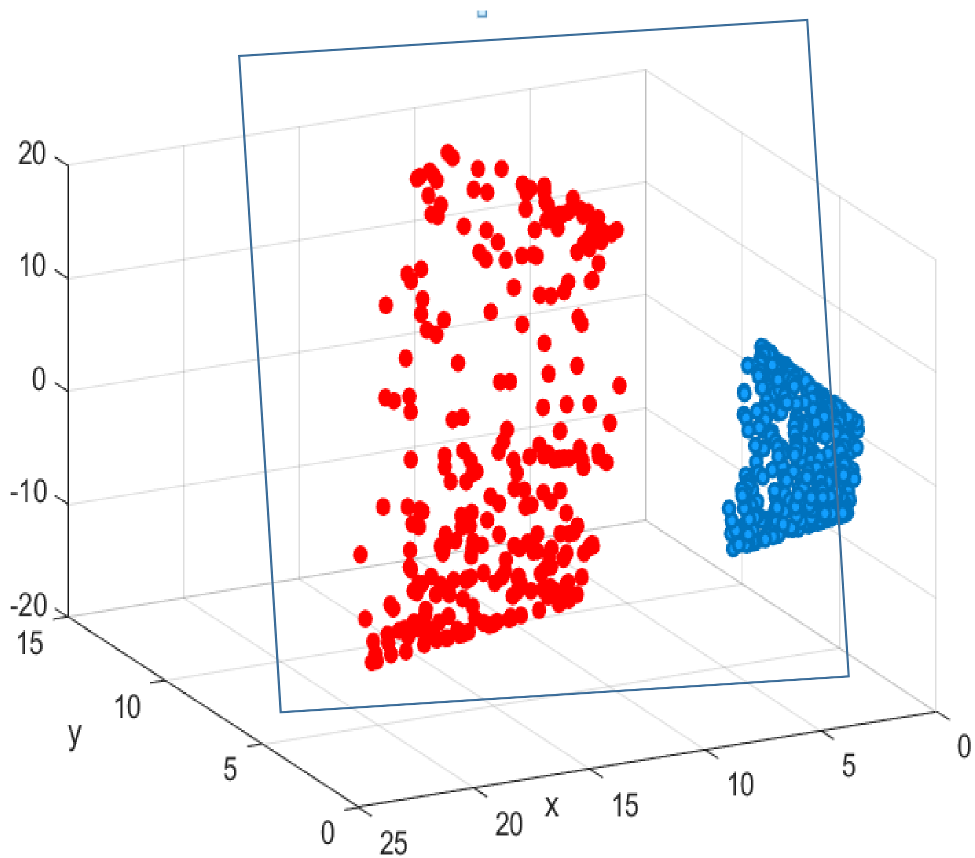


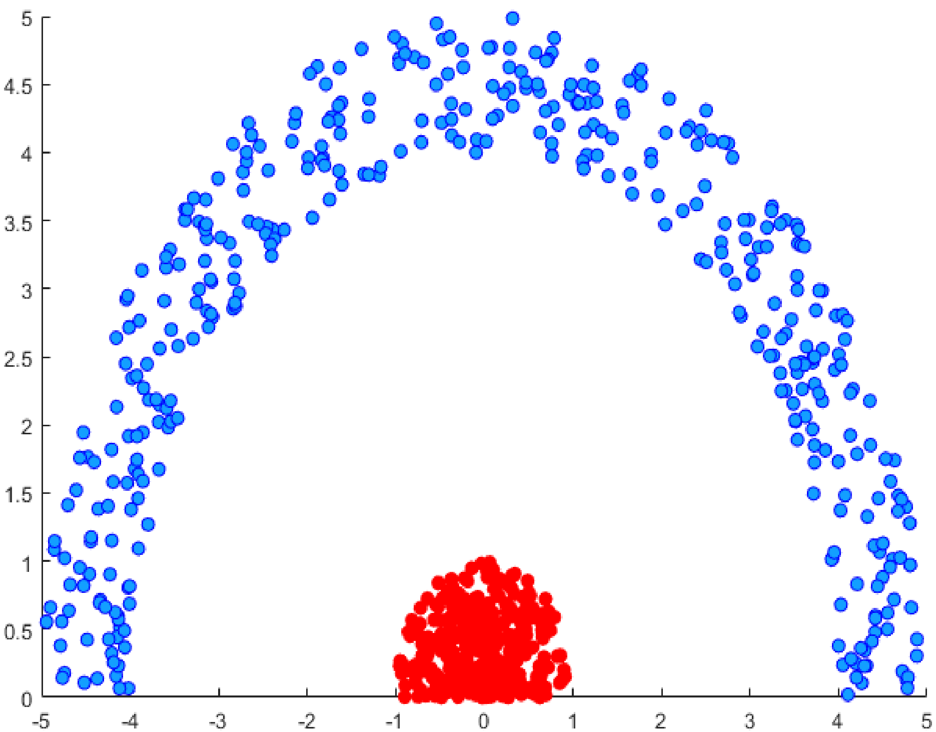
Kernel Trick

- But for different type of datasets , we require different type of mapping.

Optimization problem ??

$$\Phi(x) = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} {x_1}^2 \\ {x_2}^2 \\ \sqrt{2}x_2x_1 \end{bmatrix}$$





Non-linear SVM Optimization Problem

$$\mathrm{M}ax_{(\alpha \in R^n)} \quad -\frac{1}{2}\left(\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j\, y_i y_j \phi(x_i)^T\phi(xj)\right) \quad + \sum_{i=1}^{n}\alpha_i$$

Subject to,

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

$0 \leq \alpha_i \leq C,$ i =1,2,…C

Kernel Trick

- But for different type of datasets , we require different type of mapping.

- Thanks to mathematicians , kernel trick is available.

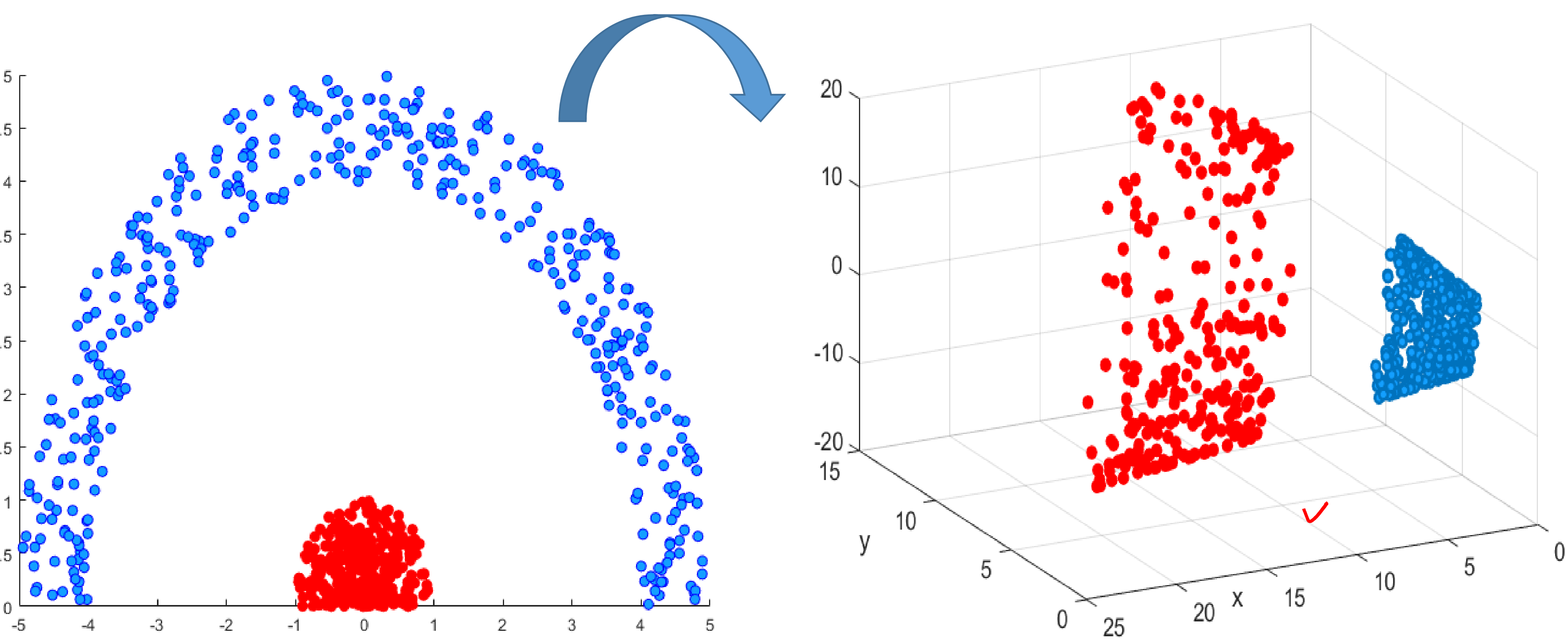
Kernel Trick

- But for different type of datasets, we require different type of corresponding mappings.

- We do not require the explicit knowledge of suitable mapping Φ for every dataset. Only the information $\Phi(x)^T\Phi(y)$ is enough.

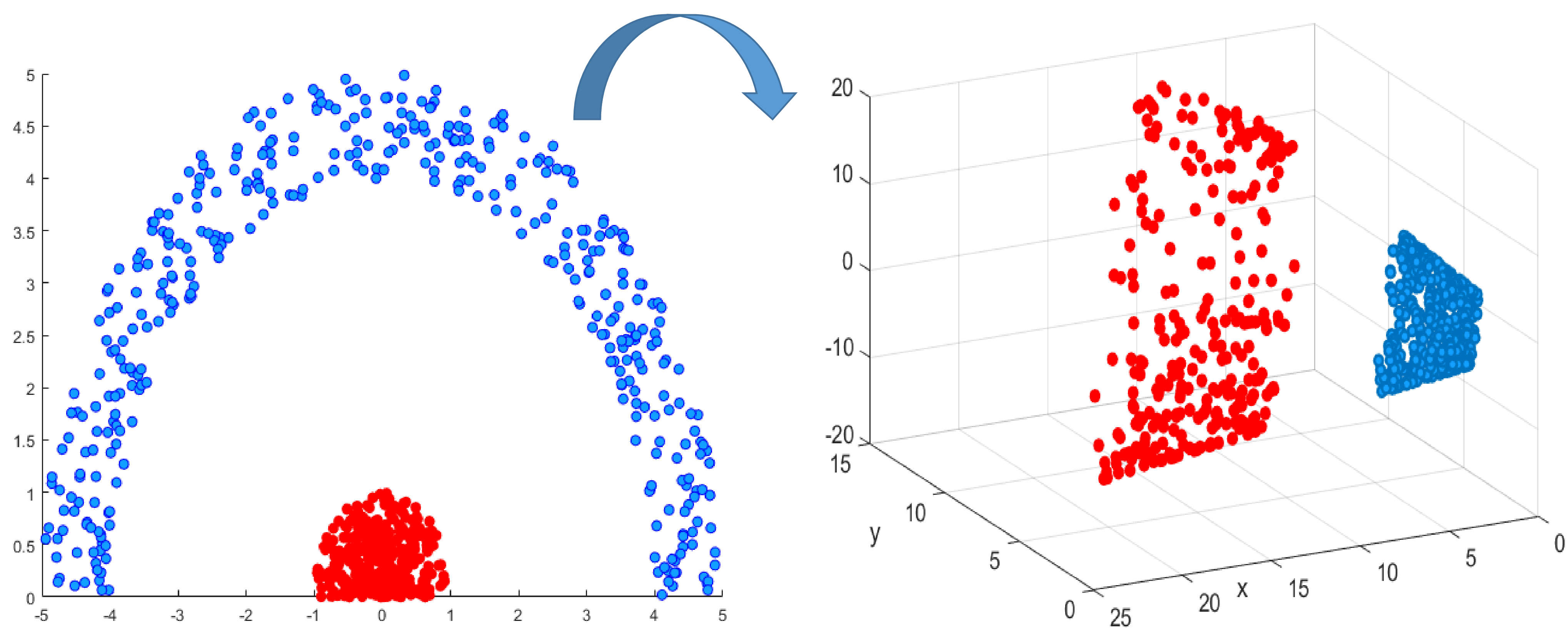
Kernel Trick

$$k(x,y) = \Phi(x)^T\Phi(y)$$



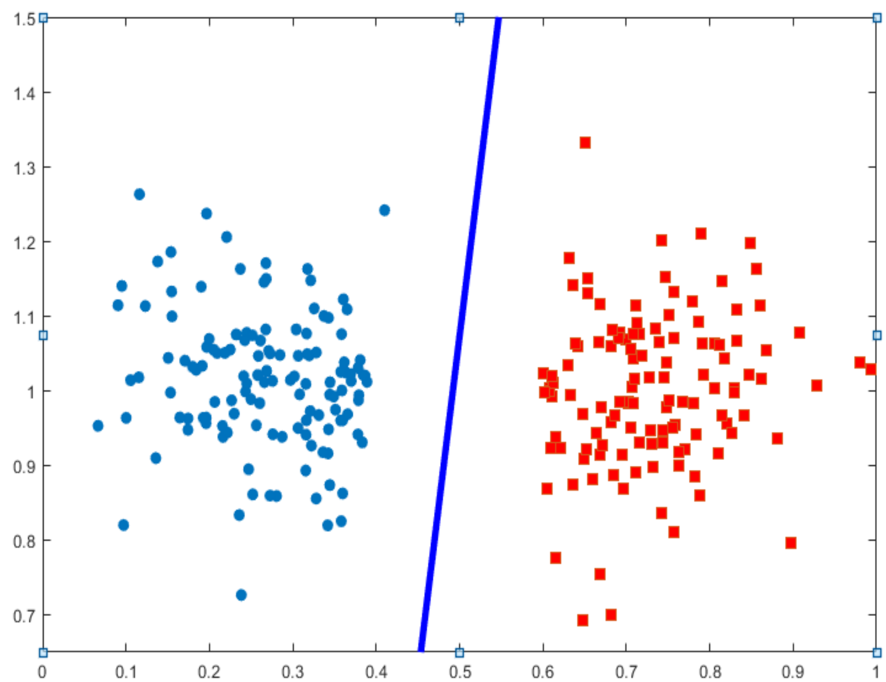
Kernel Trick

$$k(x,y) = \Phi(x)^T\Phi(y) = (x^Ty)^2$$



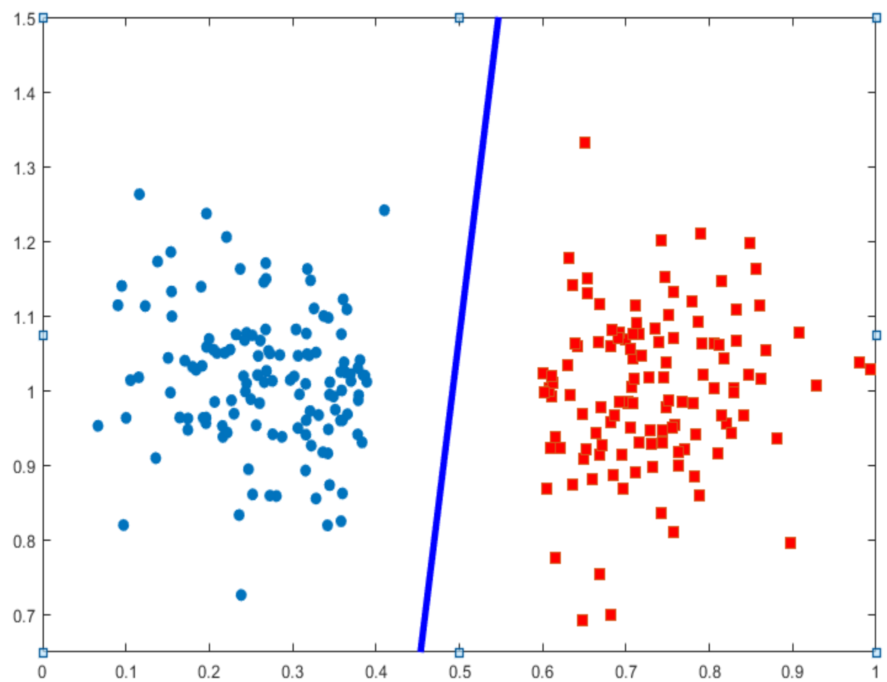
Kernel Types.

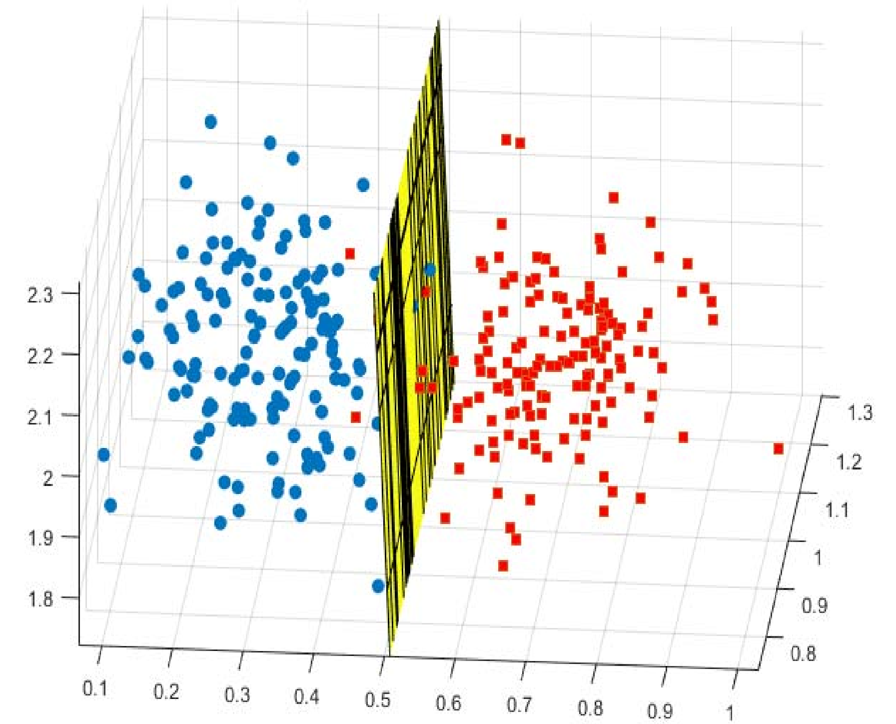
- Linear Kernel: $k(x,y)= \Phi(x)^T\Phi(y)= (x^Ty)$



Kernel Types.

- Linear Kernel: $k(x,y)= \Phi(x)^T\Phi(y)= (x^Ty)$



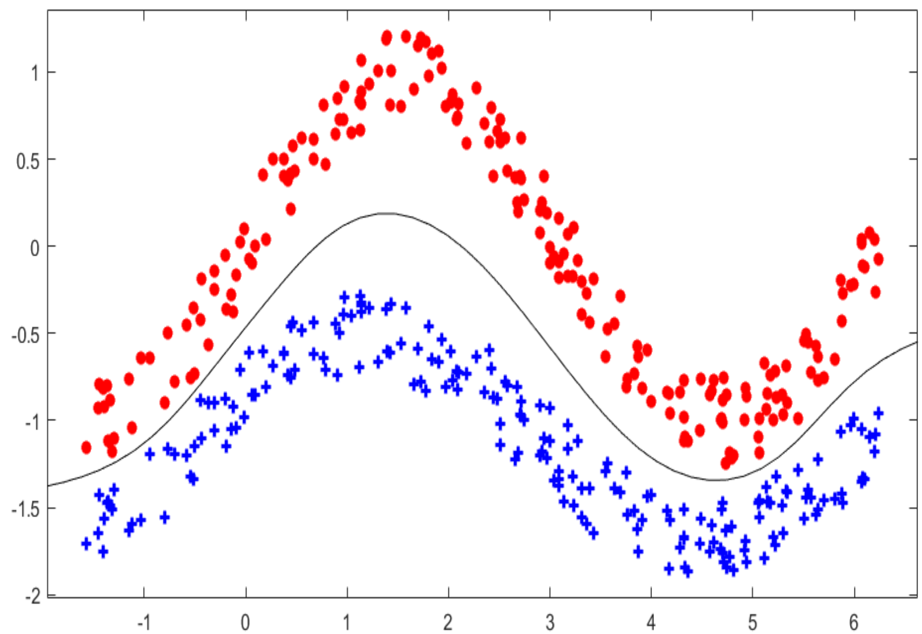


Kernel Types

- Quadratic Kernel:

$$k(x,y)= \Phi(x)^T\Phi(y)= (x^Ty+c)^2$$

,where c is the user-defined parameter.

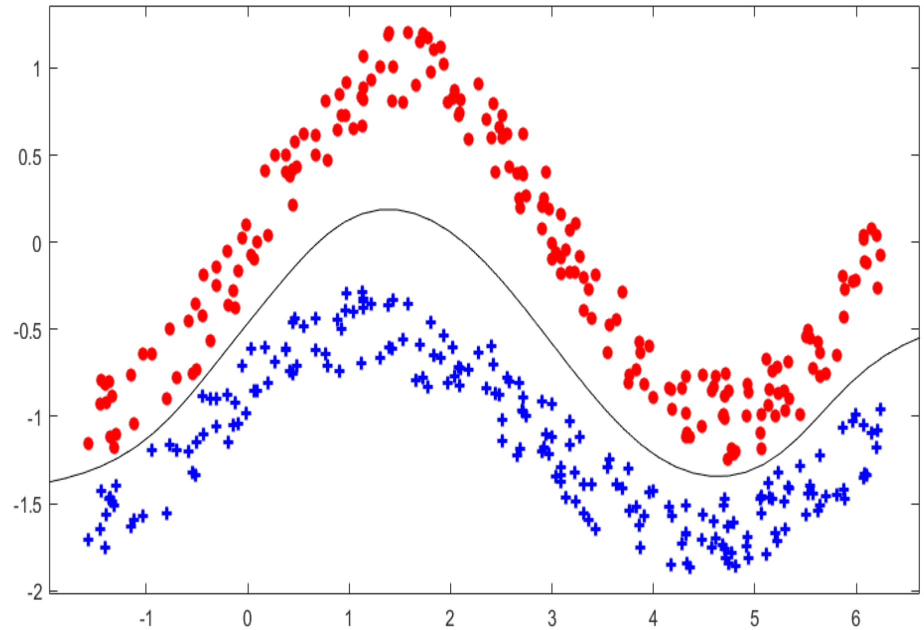


Kernel Types

- Polynomial Kernel:

$$k(x,y)= \Phi(x)^T\Phi(y)= (x^Ty+c)^p$$

,where c is the user-defined parameter.



- It can generate any type of polynomial surfaces.

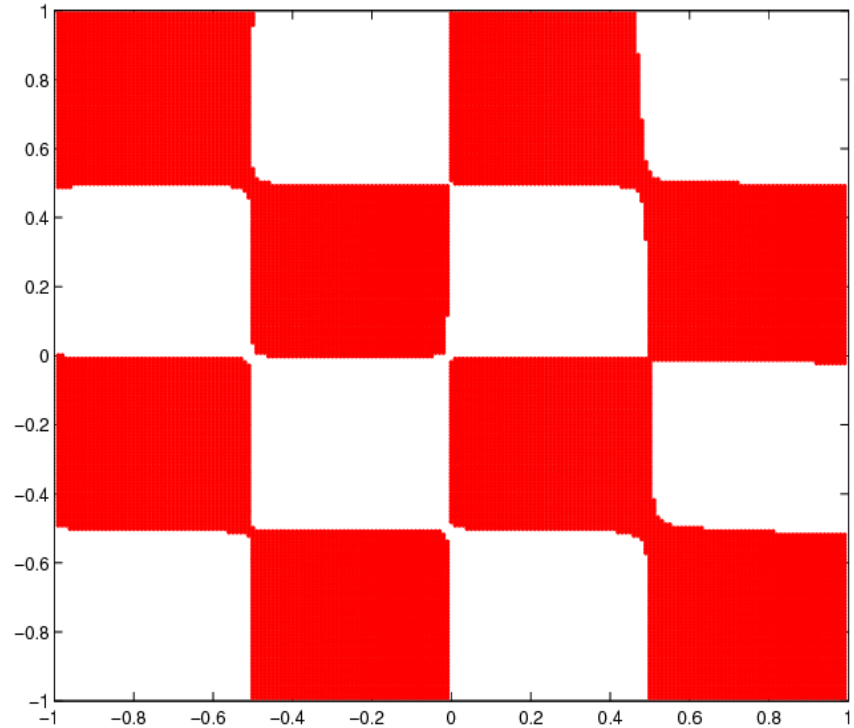
Kernel Types.

- RBF kernel:

$$k(x,y) = \Phi(x)^T\Phi(y)$$

$$= e^{-q||x-y||^2},$$

where q is the user-defined parameter.



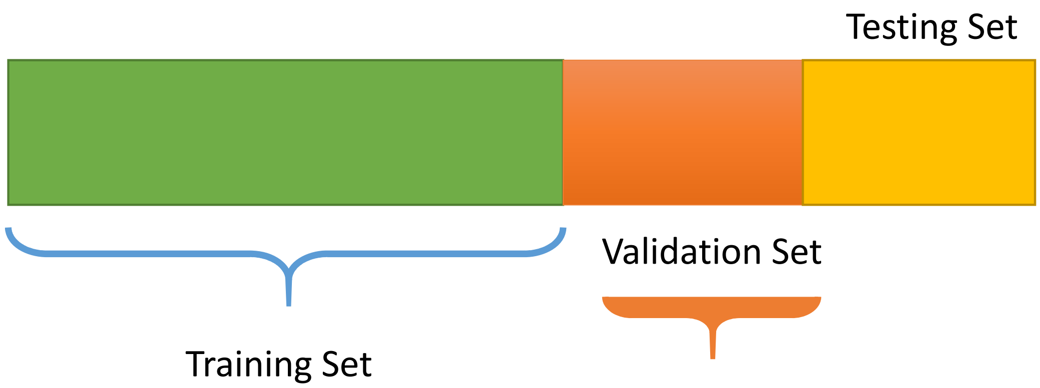
- It can generate any type of continuous surfaces.

TOY SVM

https://greitemann.dev/svm-demo

Model Selection in SVM

Model selection in SVM



Model selection in SVM

2^{-5}	2^{-4}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{4}	2^{5}	C
2^{-5}	2^{-4}	2^{-3}	2^{-2}	2^{-1}	2^{0}	2^{1}	2^{2}	2^{3}	2^{4}	2^{5}	q

Model selection in SVM

Select a value of (C,q) from the Grid

Solve the SVM optimization problem

.

Find the prediction on validation set and compute the Accuracy.

Dive Deep in SVM

Question 1

Why minimize

$$\text{Min}_{(w\in R^n, b\in R)} \quad \frac{1}{2}(w^T w) \qquad + \quad C \sum_{i=1}^{n}(Max(1 - yi(w^T x_i + b), 0))$$

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

Subject $y_i(w^T x_i + b) \geq 1$,

$i = 1, 2, \dots l.$

$$\Rightarrow (w^T x_i + b) \geq 1 \quad \text{if } y_i = 1$$

$$(w^T x_i + b) \leq -1 \quad \text{if } y_i = -1$$

$$\Rightarrow \min_{(w,b)} \frac{1}{2} w^T w$$

Subject to, $y_i(w^T x_i + b) \geq 1$,

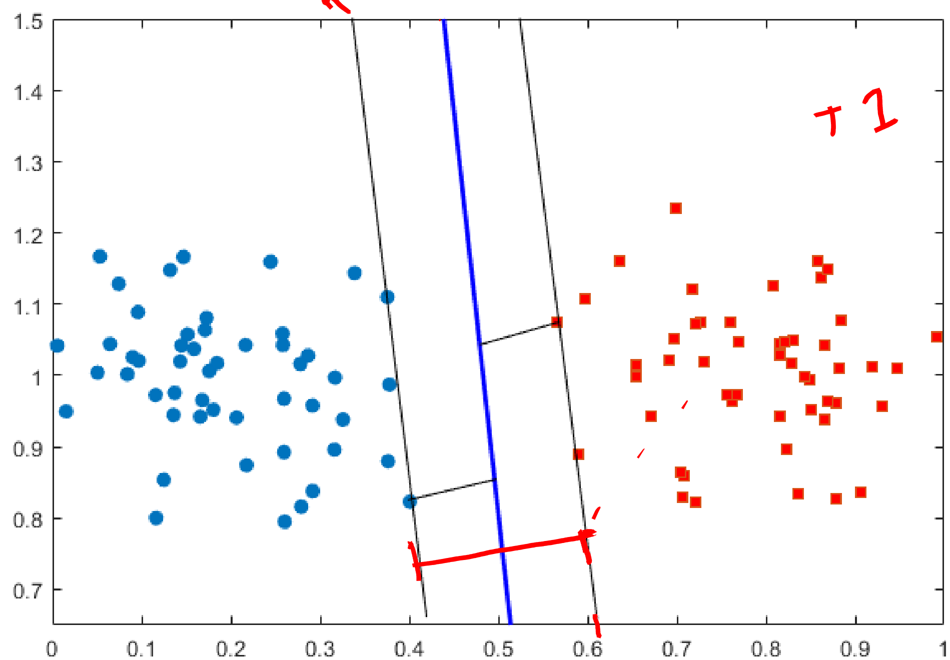
$i = 1, 2, \dots l$

$\text{sign}(w^T x + b)$

$$\frac{2}{\|w\|}$$

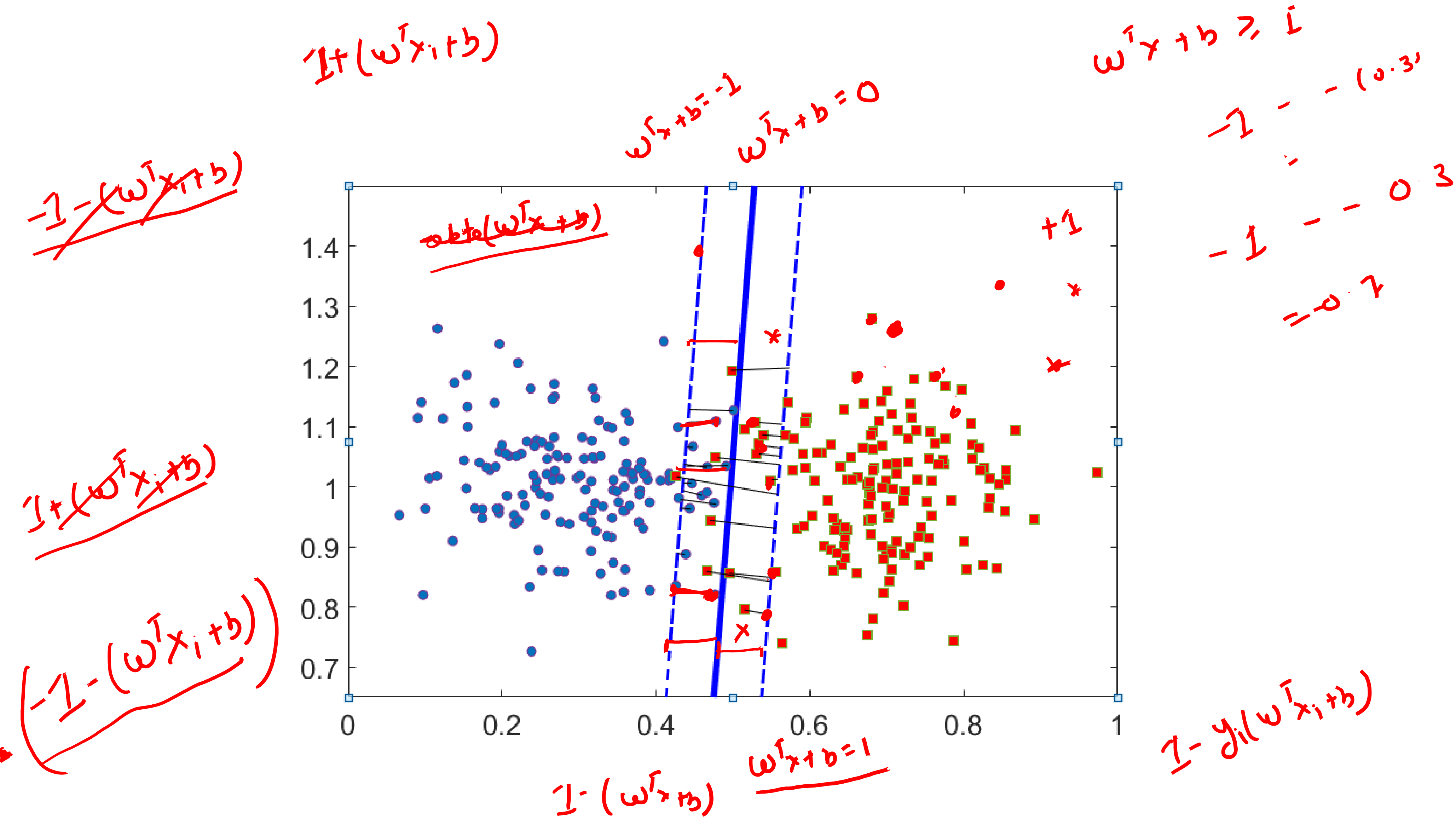
$w^T x + b = -1$

$w^T x + b = 0$



$+1$

$w^T x + b = 1$



1+(w^T x_i+b)
w^T x+b=-1
w^T x+b=0
w^T x+b ≥ 1
-1 - -(0.3)
-1 - - 0.3
=-0.7
-1-(w^T x_i+b)
+1
1+(w^T x_i+b)
-1-(w^T x_i+b)
1-(w^T x+b)
w^T x+b=1
1-y_i(w^T x_i+b)
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0
0.2
0.4
0.6
0.8
1

$$1 - y_i(w^T x_i + b)$$

$$1 - (w^T x_i + b) \quad \text{if} \quad (w^T x_i + b) \le 1, \quad y_i = 1$$

$$1 + (w^T x_i + b) \quad \text{if} \quad (w^T x_i + b) \ge -1, \quad y_i = -1$$

$$0 \quad \text{if} \quad (w^T x_i + b) \ge 1 \quad y_i = 1$$

$$0 \quad \text{if} \quad (w^T x_i + b) \le -1 \quad y_i = -1$$

$$\Rightarrow \begin{cases} 1 - y_i(w^T x_i + b) & \text{if } y_i(w^T x_i + b) \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\min_{w,b} \frac{1}{2}w^Tw + C\sum_{i=1}^{2} \max\left(1-y_i(w^Tx_i+b),\ 0\right) \quad - (1)$$

↳ 0 if $y_i(w^Tx_i+b) \geq 1$

$1-y_i(w^Tx_i+b)$, 0 therwise.

$$\xi_i = \max(1-y_i(w^Tx_i+b),\ 0)$$

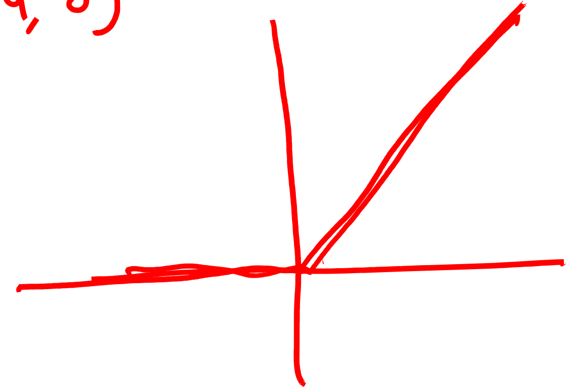
$$\min_{(w,b)} \frac{1}{2}w^Tw + C\sum_{i=1}^{2}\xi_i$$

Subject to,

$$\xi_i \geq 1-y_i(w^Tx_i+b),$$

$$\xi_i \geq 0, \quad i=1,2,\ldots 2.$$

$\max(4, 0)$ $\quad L(4)$



$d = \max(a, b)$

$d \geq a$

$d \geq b$

$w \in R^n$

$b \in R$

$\xi \in R^l$

$$\min_{(w, b, \xi)} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \quad — \quad (2)$$

Subject to,

$$y_i(w^T x_i + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots l$$

$J(w, b)$

$$\min_{(w, b)} \underbrace{\frac{1}{2} w^T w}_{w} + C \sum_{i=1}^{l} \underbrace{\max(0, \; 1 - y_i(w^T x_i + b))}_{} \quad — \quad (1)$$

$w \quad +$

$$\nabla_w J(w,b) = w + C \sum_{i=1}^{l} \begin{cases} 0 & \text{if } y_i(w^T x_i + b) \geq 1 \\ -y_i x_i & \text{if } y_i(w^T x_i + b) < 1 \end{cases}$$

$$\nabla_b J(w,b) = C \sum_{i=1}^{l} \begin{cases} 0 & \text{if } y_i(w^T x_i + b) \geq 1 \\ -y_i & \text{if } y_i(w^T x_i + b) < 1 \end{cases}$$

Intialize $\overset{0}{w} \in R^n$, $\overset{0}{b} \in R$

$$w^{k+1} = w^k - \eta \nabla_w J(w,b)$$

$$b^{k+1} = b^k - \eta \nabla_b J(w,b)$$

until $\left\| \begin{matrix} \nabla_w J(w^{k+1}, b^{k+1}) \\ \nabla_b J(w^{k+1}, b^{k+1}) \end{matrix} \right\| \leq \epsilon$

$$\min_{w, b} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \underbrace{\max(1 - y_i(w^T \phi(x_i) + b), 0)}$$

Decision Boundary $\quad \text{sign}(w^T \phi(x) + b)$

$$\phi(x_i)^T \phi(x_j) = K(x_i, x_j)$$

$$w^T \phi(x) + b$$

$$\cancel{w} = \left(\sum_{i=1}^{l} u_i \phi(x_i) \right)^T \phi(x) + b \quad = \quad \underbrace{\sum_{j=1}^{l} K(x_i, x) u_i + b}$$

u^T

$$\min_{\substack{u,b \\ u\in R^{l} \\ b\in R}}^{J(u,b)} \frac{1}{2}u^T u + C\sum_{i=1}^{l} \max\left(1 - y_i\left(\sum_{j=1}^{l} K(x_j, x)u_j + b\right), 0\right)$$

$$\nabla_u J(u,b) = u + C\sum_{i=1}^{l} \begin{cases} 0 & \text{if } y_i\left(\sum_{j=1}^{l} K(x_j,x)u_j + b\right) > 1 \\ -y_i \begin{bmatrix} K(x_i, x_1) \\ K(x_i, x_2) \\ \vdots \\ K(x_i, x_l) \end{bmatrix}, & \text{otherwise} \end{cases}$$

x_i

x_i, x_1

x_1

$$= 0 + C\sum_{i=} \begin{cases} 0 \\ -y_i \end{cases}$$

$$\min_{w,b,\xi} \left\{ \min_{\substack{w\in\mathbb{R}^n \\ b\in\mathbb{R}}} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \right.$$

Subject to,

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

$$\left. \xi_i \geq 0 \quad i = 1, 2, \ldots, l \right\}$$

~~Man~~

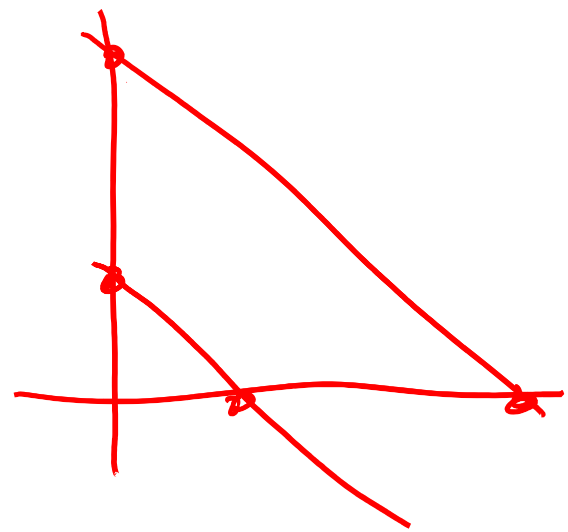
$$\min_{x_1, x_2} \; 3x_1 + 4x_2$$

Subject to,

$$5x_1 + 6x_2 \leq 5$$

$$6x_1 + 9x_2 \leq 9$$

$$x_1, x_2 \geq 0$$



$$\min_{x_1, x_2} \quad 5x_1^2 + 6x_1x_2 + 3x_2^2$$

Subject to,

$$4x_1 + 3x_2 \leq 5$$

$$5x_1 + 6x_2 \leq 9$$

$$x_1, \; x_2 \geq 0$$

$$\left\{ \begin{array}{l} \min\limits_{x \in \mathbb{R}^n} \quad f(x) \\ \text{Subject to,} \\ \quad g_i(x) \leq 0 \;, \; i = 1, 2, \ldots m \\ \quad h_j(x) = 0 \;, \; j = 1, 2, \ldots k. \end{array} \right.$$

$$\min_{x,\, y} \quad x^2 + y^2$$

subject to

$$f(x)$$

$$\underbrace{x^2 + 2xy}_{} \leq 5$$

$$x^2 \leq 5$$

$$\min_{x \in \mathbb{R}^n} \underbrace{L(x, \alpha, \beta)}_{\cancel{x \in D}} = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{j=1}^{k} \beta_j h_j(x)$$

$$D = \Big\{ x \in \mathbb{R}^n : g_i(x) \leq 0, \; i = 1, 2, \ldots m$$
$$h_j(x) = 0, \; j = 1, 2, \ldots k \Big\}$$

$$\underset{x \in D}{f(x)} \geq \underset{x \in D}{L(x, \alpha, \beta)}$$

$$\inf_{x \in D} f(x) \geq \inf_{x \in D} L(x, \alpha, \beta)$$

$$\inf_{x \in D} f(x) \geq \inf_{x \in D} L(x, \alpha, \beta) \geq \inf_{x \in R^n} L(x, \alpha, \beta)$$

$$\Rightarrow \boxed{\inf_{x \in D} f(x)} \geq \underbrace{\inf_{x \in R^n} L(x, \alpha, \beta)}$$

$$\max_{\alpha,\beta} \left(\inf_{x \in \mathbb{R}^n} L(x,\alpha,\beta) \right) \quad \infty \qquad p^*$$

$$= \inf_{x \in D} f(x)$$

$$\min_{x \in D} f(x) \Rightarrow \max_{\alpha,\beta} \inf \qquad = p^*$$

$$\max_{\alpha,\beta} \left(\inf_{x \in \mathbb{R}^n} L(x,\alpha,\beta) \right)$$

$$L(x,\alpha,\beta) = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{j=1}^{k} \beta_j h_j(x)$$

KKT conditions

Consider the convex programming problem (i), the point x^* is said to satisfy the KKT condition, if there exists multipliers $(\alpha_1^*, \alpha_2^*, \alpha_m^*)$ and $(\beta_1^*, \beta_2^*, \ldots \beta_k^*)$ such that Lagrangian fun^c satisfies

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{j=1}^{k} \beta_j h_j(x)$$

Satisfies

$$\alpha_i g_i(x) = 0$$

$f(x)$ $g_i(x^*) \leq 0$; $i = 1, 2, \ldots m$

$h_i(x^*) = 0$, $i = 1, 2, \ldots k$

$\alpha_i, \beta_i \geq 0$

$\alpha_i g_i(x) = 0$

$\beta_i h_i(x) = 0$

$$\nabla_x L(x, \alpha^*, \beta^*) = \nabla f(x^*) + \sum_{i=1}^{m} \nabla \alpha_i g_i(x) + \sum_{j=1}^{k} \nabla \beta_i h_i(x) = 0$$

if x^* satisfies all condition then it is solution

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$

subject to

$$g_i(x) \leq 0, \quad i = 1, 2, \ldots m$$

$$h_j(x) = 0; \quad j = 1, 2, \ldots K$$

$$\max_{\alpha, \beta} \left(\inf_{x \in \mathbb{R}^n} L(x, \alpha, \beta) \right)$$

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{j=1}^{K} \beta_j h_j(x)$$

$$\nabla_x L(x^*, \alpha, \beta) = \nabla f(x^*) + \sum_{i=1}^{m} \alpha_i \nabla g_i(x^*) + \sum_{j=1}^{k} \beta_j \nabla h_j(x^*) = 0$$

$$g_i(x^*) \leq 0 \qquad i = 1, 2, \ldots m$$

$$h_j(x^*) = 0 \qquad j = 1, 2 \ldots k$$

$$\alpha_i \geq 0 \qquad i = 1, 2, \ldots m$$

$$\beta_j \geq 0 \qquad j = 1, 2, \ldots k.$$

$$\alpha_i g_i(x) = 0, \quad i = 1, 2, \ldots m$$

$$\beta_j h_j(x) = 0, \quad j = 1, 2 \ldots k.$$

$$\min_{\substack{w \in \mathbb{R}^n \\ b \in \mathbb{R} \\ \xi \in \mathbb{R}^l}} \; \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

$w^T \varphi(x_i)$

$w^T \varphi(x_i)$ — 1

Subject to

$y_i(w^T x_i + b) \geq 1 - \xi_i \quad , \quad i = 1, 2 \ldots l$

$\xi_i \geq 0 \quad , \quad i = 1, 2, \ldots l$

$$\boxed{-(y_i(w^T x_i + b) - 1 + \xi_i)} \leq 0 \quad i = 1, 2 \ldots l$$

$\varphi(x_i)$

$-\xi_i \leq 0 \quad , \quad i = 1, 2 \ldots l$

$L(\cancel{x}, \alpha, \beta$

$$\xi_1 + \xi_2 + \cdots \xi_l$$

$$\inf_{w,b}\left(L(w, b, \xi, \alpha, \beta) \right) = \cancel{f} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

$$- \sum_{i=1}^{l} \alpha_i \left(y_i (w^T \underset{\phi(x_i)}{x_i} + b) - 1 + \xi_i \right)$$

$$w = \sum_{i=1}^{l} \alpha_i y_i \underset{\phi(x_i)}{x_i}$$

$$- \sum_{j=1}^{l} \beta_i \xi_i$$

$$\Rightarrow \left\{ \max_{\alpha, \beta} \left(\inf_{w,b} \left(L(w, b, \xi, \alpha, \beta) \right) \right) \right\}$$

$$\nabla_w L(w, b, \xi, \alpha, \beta) = 0$$

$$\Rightarrow \quad w - \sum_{i=1}^{l} \alpha_i y_i \overset{\phi(x_i)}{x_i} = 0$$

$$\Rightarrow \quad w = \sum_{i=1}^{l} \alpha_i y_i \overset{\phi(x_i)}{x_i} \quad \text{— (1)}$$

$$\nabla_b L(w, b, \xi, \alpha, \beta) = 0 \quad \Rightarrow \quad \sum_{i=1}^{l} \alpha_i \underline{y_i} = 0 \quad \text{— (2)}$$

$$\nabla_\xi L(w, b, \xi, \alpha, \beta) = 0$$

$$\Rightarrow \quad C - \alpha_i - \beta_i = 0, \quad i = 1, 2, \dots, \underline{l} \quad \text{— (3)}$$

$$\max_{\alpha,\beta}\left(\inf_{w,b,\xi} L(w, b, \xi, \alpha, \beta)\right)$$

$$\sum_{i=1} \alpha_i y_i b$$

$$\max_{\alpha,\beta}\left(\inf_{w,b,\xi}\left(\frac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i \left(y_i(w^T x_i + b) - 1 + \xi_i\right) - \sum_{i=1}^{l} \beta_i \xi_i\right)\right)$$

$\phi(x_i)$

$$\max_{\alpha,\beta} \frac{1}{2}\left(\sum_{i=1}^{l} \alpha_i y_i x_i\right)^T\left(\sum_{j=1}^{l} \alpha_j y_j x_j\right) - \sum_{i=1}^{l} \alpha_i \left(y_i\left(\left(\sum_{j=1}^{l} \alpha_j y_j x_j\right)^T x_i\right) - 1\right)$$

$\phi(x_i)$ $\phi(x_j)$ $\phi(x_j)$

$$\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$\phi(x_i)$

Σ

subject to, $\sum_{i=1} \alpha_i y_i = 0$, $C - \alpha_i - \beta_i = 0$, $i = 1, 2, \ldots l$

$$\max_{\alpha,\beta} \; -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j (x_i^T x_j) + \sum_{i=1}^{l}\alpha_i$$

$$\phi(x_i)^T\phi(x_j)$$

$C - \alpha_i \geq 0$

$0 < \alpha_i < C$

Subject to,

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

$$C - \alpha_i - \beta_i = 0, \quad i = 1, 2, \ldots l.$$

$$\alpha_i, \beta_i \geq 0, \quad i = 1, 2, \ldots l$$

$$\min_{\alpha,\beta} \; \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j (x_i^T x_j) - \sum_{i=1}^{l}\alpha_i$$

$$\phi(x_i)^T\phi(x_j) \qquad K(x_i, x_j)$$

Subject to,

$$\sum_{i=1}^{l}\alpha_i y_i = 0, \qquad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots l$$

$$\max_{\alpha, \beta} \quad \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j \beta_i \beta_j K(x_i, x_j) - \sum_{i=1}^{l} \alpha_i$$

Subject to, $\displaystyle \sum_{i=1}^{l} \alpha_i y_i = 0$

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, l$$

$$w = \sum_{i=1}^{l} \alpha_i y_i \varphi(x_i)$$

$$w^T \varphi(x_j) = \left(\sum_{i=1}^{l} \alpha_i y_i \varphi(x_i) \right)^T \varphi(x_j) = \sum_{i=1}^{l} \alpha_i y_i K(x_i, x_j)$$

(w, b, ξ) will be optimal soln of problem ① if it satisfies KKT conditions

$\alpha_i g_i(x) = 0$

$$\nabla_w L(w, b, \xi, \alpha, \beta) = 0 \Rightarrow w^* = \sum_{i=1}^{l} \alpha_i y_i \phi(x_i)$$

$$\nabla_b L(w, b, \xi, \alpha, \beta) = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$$

$$\nabla_\xi L(b, \xi, \alpha, \beta) \Rightarrow C - \alpha_i - \beta_i = 0 \Rightarrow 0 \leq \alpha_i \leq C$$

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, l$$

$$\xi_i \geq 0 \quad i = 1, 2, \dots, l$$

$$\alpha_i(y_i(w^T \phi(x_i) + b) - 1 + \xi_i) = 0, \quad i = 1, 2, \dots, l$$

$$\beta_i \xi_i = 0, \quad i = 1, 2, \dots, l, \qquad \beta_i \geq 0 \qquad \alpha_i \geq 0$$

y_j

Consider a point (x_j, y_j) such that

$$y_j(w^T x_j + b) \geq 1 \Rightarrow \xi_j = 0$$

$$\alpha_j \neq 0 \Rightarrow \underline{y_j(w^T\phi(x_j) + b) - 1 + \xi_j} = 0$$

$\Rightarrow$

$$\alpha_j < C \Rightarrow \beta_j > 0 \Rightarrow \xi_j = 0$$

$\Rightarrow$ if $0 < \alpha_j < C$

$$\underline{y_j(w^T\phi(x_j) + b) - 1} = 0$$

$\cancel{y_j b = y_j - w^T\phi(x_j)}$ $\Rightarrow$ $y_j(w^T\phi(x_j)) + y_j b = 1$

$$\frac{1}{y_j} - \cancel{y_j}\,(w^T\phi(x_j)) = b$$

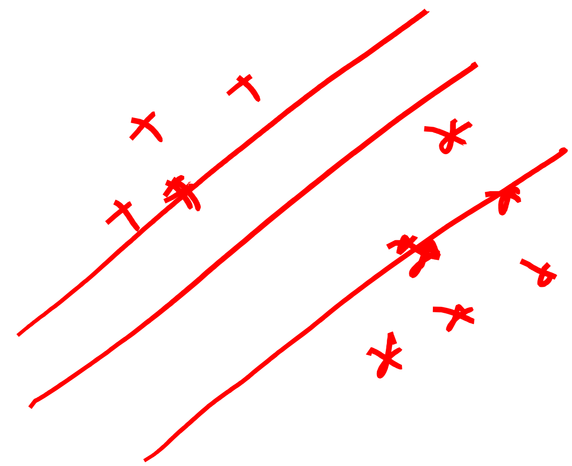
$$y_j \cdot \frac{1}{1}$$

$$\frac{1}{y_j} = y_j$$

$$\Rightarrow b = \frac{1}{y_j} - w^T\phi(x_j)$$

$$= y_j - w^T\phi(x_j)$$

$$= y_j - \sum_{i=1}^{l} \alpha_i y_i K(x_i, x_j)$$



$$0 < \alpha_i < C$$