Applied Text Analytics & Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Support Vector Machine – Part 2

Now We Need to Maximize the Margin

Maximize $\frac{2}{||\theta||}$

Subject to Min value of $|x_i\theta + b| = 1 \Rightarrow nearest\ neighbour$

$i = 1,2, \dots, N$

There is a "min" in our constraining; it can be hard to optimize this problem (non-convex form)

Can I write the following term to get rid of absolute value?

$|x_i\theta + b| = y_i(x_i\theta + b) \Rightarrow$ for a correct classification

If min $|x_i\theta + b| = 1 \Rightarrow so\ it\ can\ be\ at\ least\ 1$

Maximize $\frac{2}{||\theta||}$

Subject to $y_i(x_i\theta + b) \geq 1$ for $i = 1,2, \dots, N$

Now We Need to Maximize the Margin

Maximize $\frac{2}{||\theta||}$

Subject to Min value of $|x_i\theta + b| = 1 \Rightarrow nearest\ neighbour$

$i = 1,2,\dots,N$

There is a "min" in our constraining; it can be hard to optimize this problem (non-convex form)

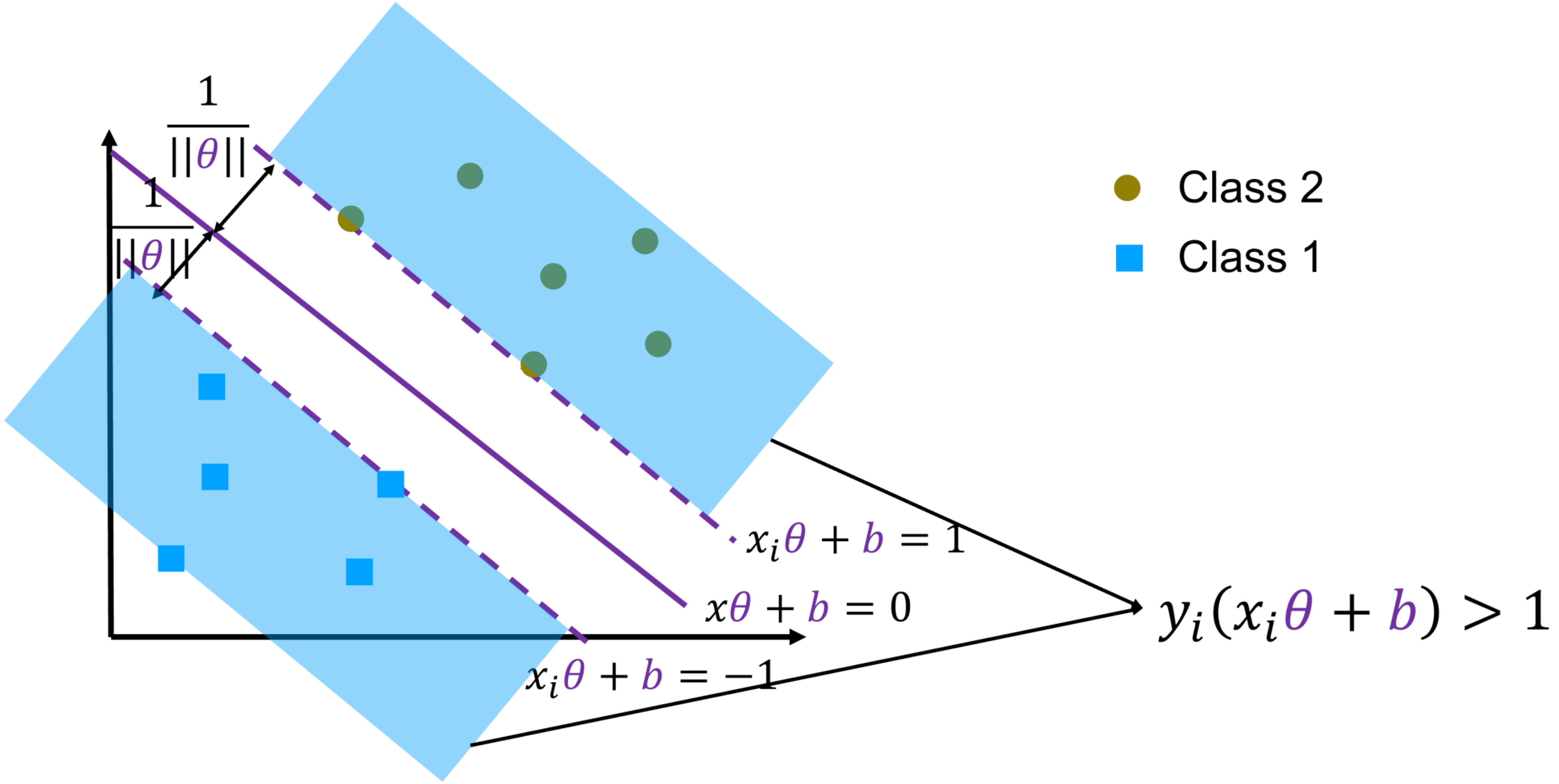
Can I write the following term to get rid of absolute value?

$|x_i\theta + b| = y_i(x_i\theta + b) \Rightarrow$ for a correct classification

If min $|x_i\theta + b| = 1 \Rightarrow so\ it\ can\ be\ at\ least\ 1$

Maximize $\frac{2}{||\theta||}$

Subject to $y_i(x_i\theta + b) \geq 1$ for $i = 1,2,\dots,N$

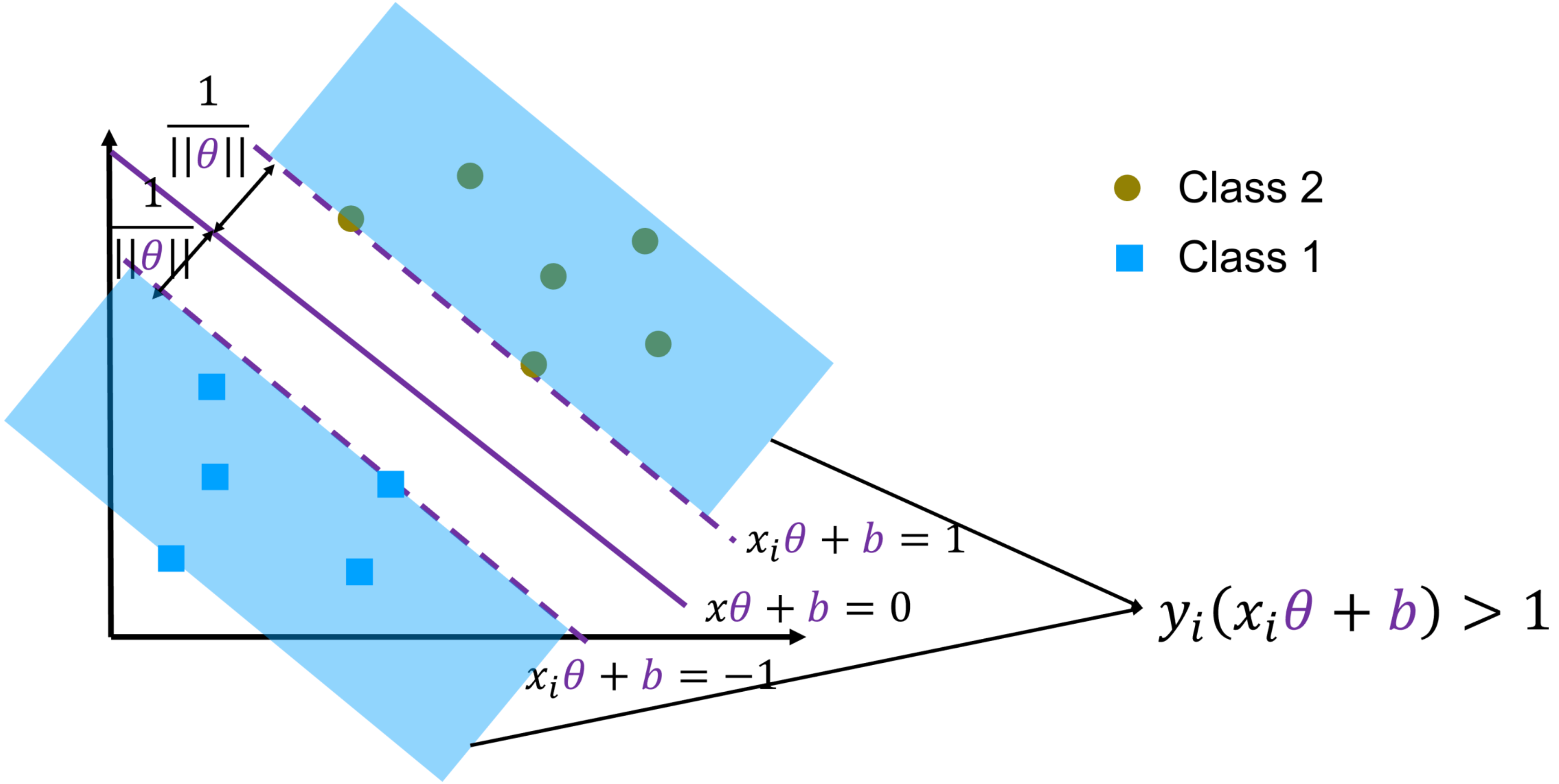


Maximize $\frac{2}{||\theta||}$

Subject to $y_i(x_i\theta + b) \geq 1$ for $i = 1,2,\dots,N$

Minimize $\frac{1}{2}\theta\theta^T$

Subject to $y_i(x_i\theta + b) \geq 1$ for $i = 1,2,\dots,N$



Maximize $\frac{2}{||\theta||}$

Subject to $y_i(x_i\theta + b) \geq 1$ for $i = 1,2,\ldots,N$

Minimize $\frac{1}{2}\theta\theta^T$

Subject to $y_i(x_i\theta + b) \geq 1$ for $i = 1,2,\ldots,N$

Lagrange Formulation

Minimize $\frac{1}{2}\theta\theta^T$ s.t. $y_i(x_i\theta + b) - 1 \geq 0$

$$\mathcal{L}(\theta, b, \alpha) = \frac{1}{2}\theta\theta^T - \sum_{i=1}^{N} \alpha_i(y_i(x_i\theta + b) - 1)$$

Minimize w.r.t θ and b and maximize *w.r.t* each $\alpha_i \geq 0$

$$\nabla_\theta \, \mathcal{L}(\theta, b, \alpha) = \theta - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$

$$\nabla_b \, \mathcal{L}(\theta, b, \alpha) = -\sum_{i=1}^{N} \alpha_i y_i = 0$$

Lagrange Formulation

Minimize $\frac{1}{2}\theta\theta^T$ s.t. $y_i(x_i\theta + b) - 1 \geq 0$

$$\mathcal{L}(\theta, b, \alpha) = \frac{1}{2}\theta\theta^T - \sum_{i=1}^{N} \alpha_i(y_i(x_i\theta + b) - 1)$$

Minimize w.r.t θ and b and maximize *w.r.t* each $\alpha_i \geq 0$

$$\nabla_\theta\, \mathcal{L}(\theta, b, \alpha) = \theta - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$

$$\nabla_b\, \mathcal{L}(\theta, b, \alpha) = -\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\theta = \sum_{i=1}^{N} \alpha_i y_i x_i \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\mathcal{L}(\theta, b, \alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\theta\theta^T$$

$$\mathcal{L}(\theta, b, \alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i x_j^T$$

maximize *w.r.t* each $\alpha_i \geq 0$ for $i = 1, \ldots, N$

and

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

$$\theta = \sum_{i=1}^{N} \alpha_i y_i x_i \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

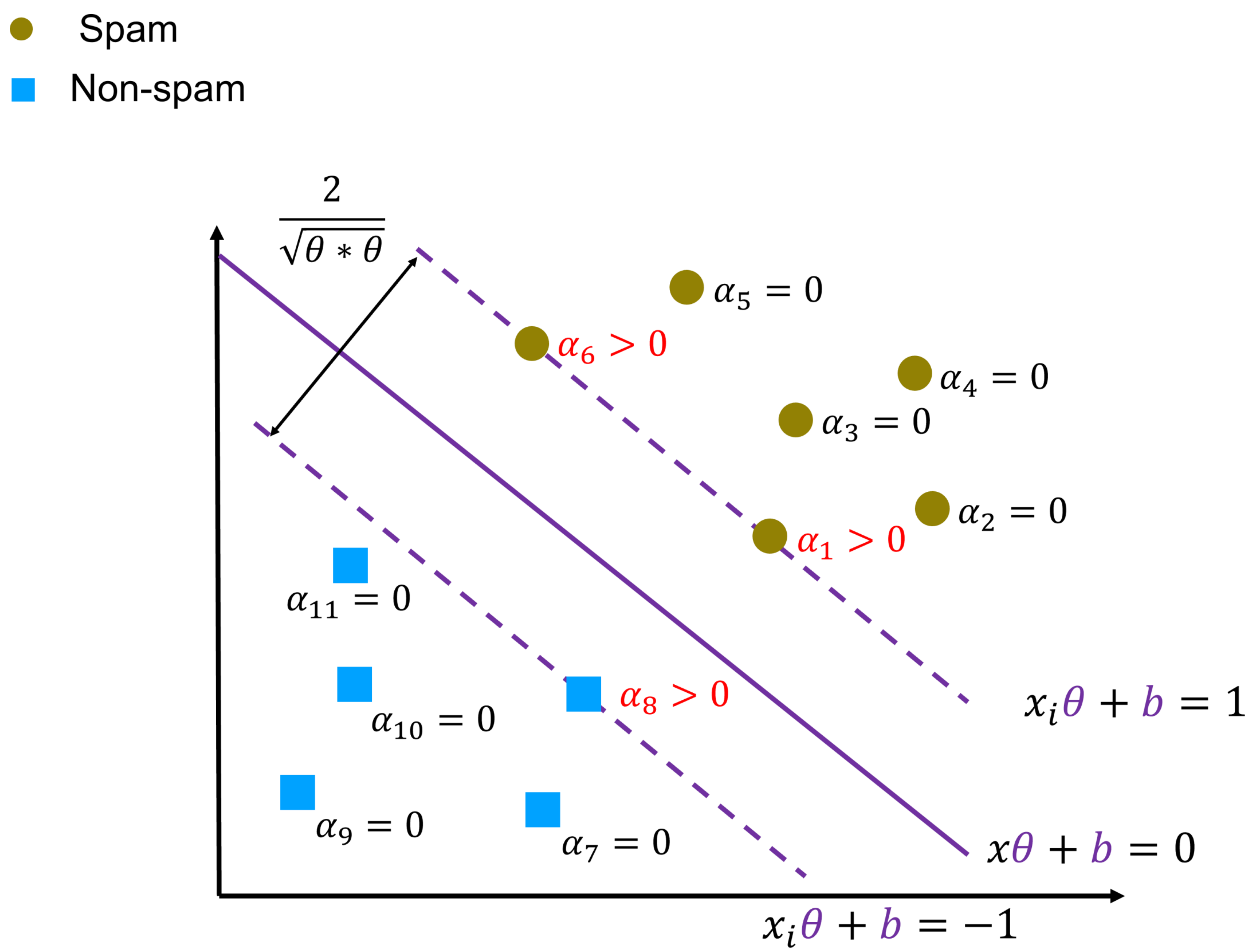
$$\mathcal{L}(\theta, b, \alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \theta \theta^T$$

$$\mathcal{L}(\theta, b, \alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i x_j^T$$

maximize *w.r.t* each $\alpha_i \geq 0$ for $i = 1, \dots, N$

and

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



Spam
Non-spam
$\frac{2}{\sqrt{\theta * \theta}}$
$\alpha_5 = 0$
$\alpha_6 > 0$
$\alpha_4 = 0$
$\alpha_3 = 0$
$\alpha_2 = 0$
$\alpha_1 > 0$
$\alpha_{11} = 0$
$\alpha_{10} = 0$
$\alpha_8 > 0$
$\alpha_9 = 0$
$\alpha_7 = 0$
$x_i\theta + b = 1$
$x\theta + b = 0$
$x_i\theta + b = -1$

Training

$$\theta = \sum_{i=1}^{N} \alpha_i y_i x_i$$

No need to go over all datapoints

$$\rightarrow \theta = \sum_{x_i\, in\, SV} \alpha_i y_i x_i$$

and for b pick any support vector and calculate: $y_i(x_i\theta + b) = 1$

Testing

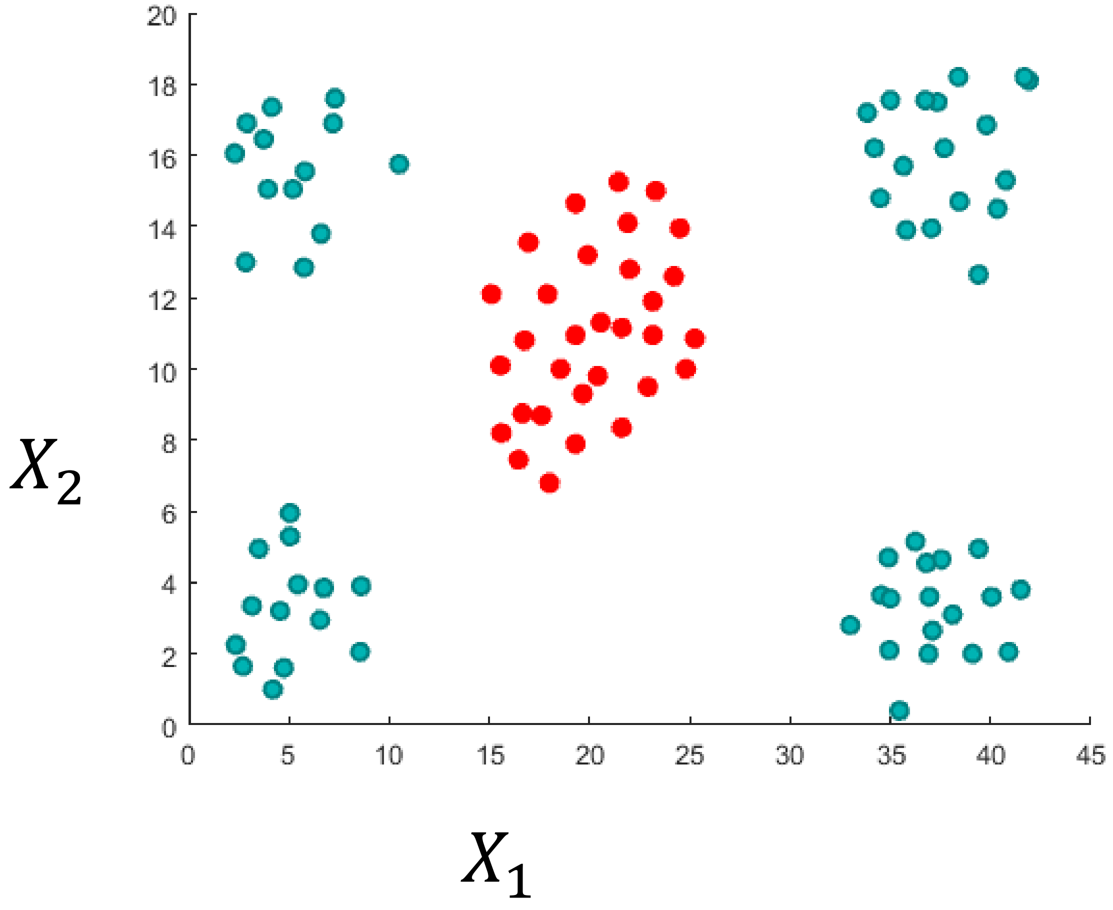
For a new test point s

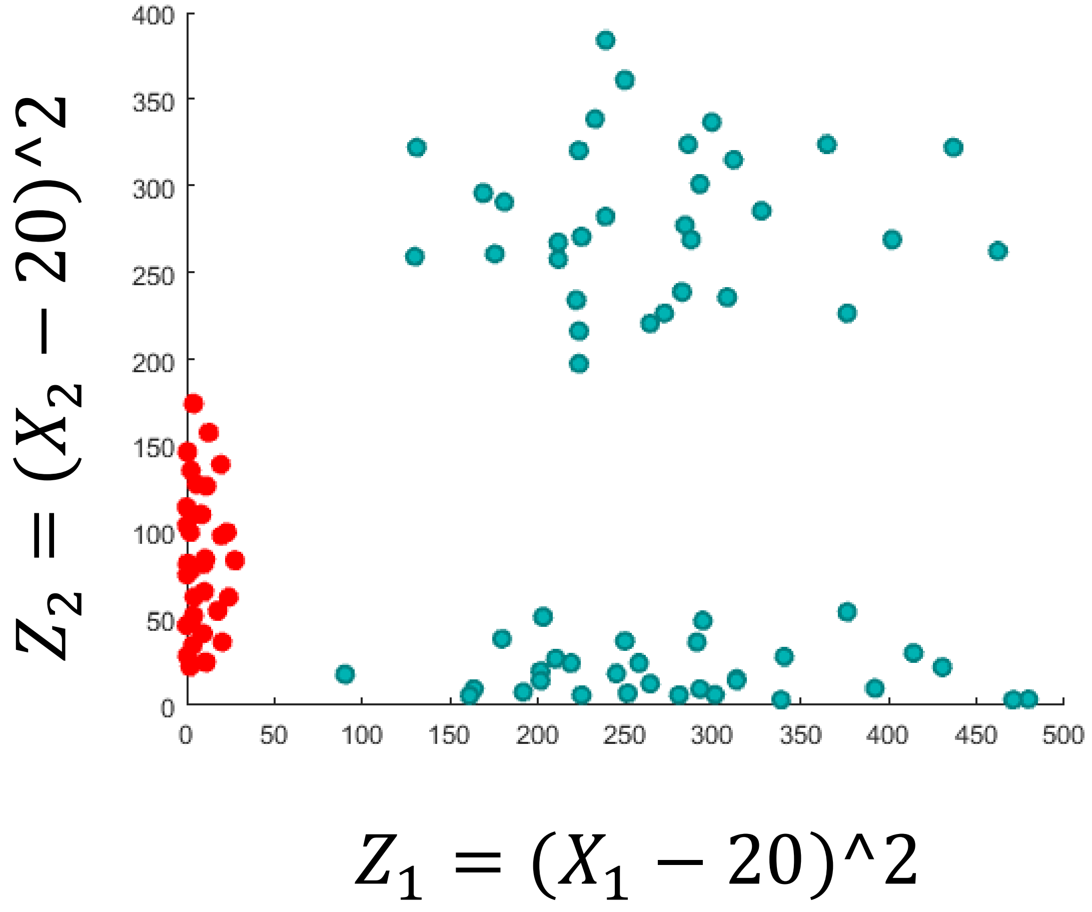
Compute:

$$\mathrm{s}\theta + \mathrm{b} = \sum_{x_i\, in\, SV} \alpha_i y_i x_i s^T + b$$

Classify s as class 1 if the result is positive, and class 2 otherwise

From x to z Space

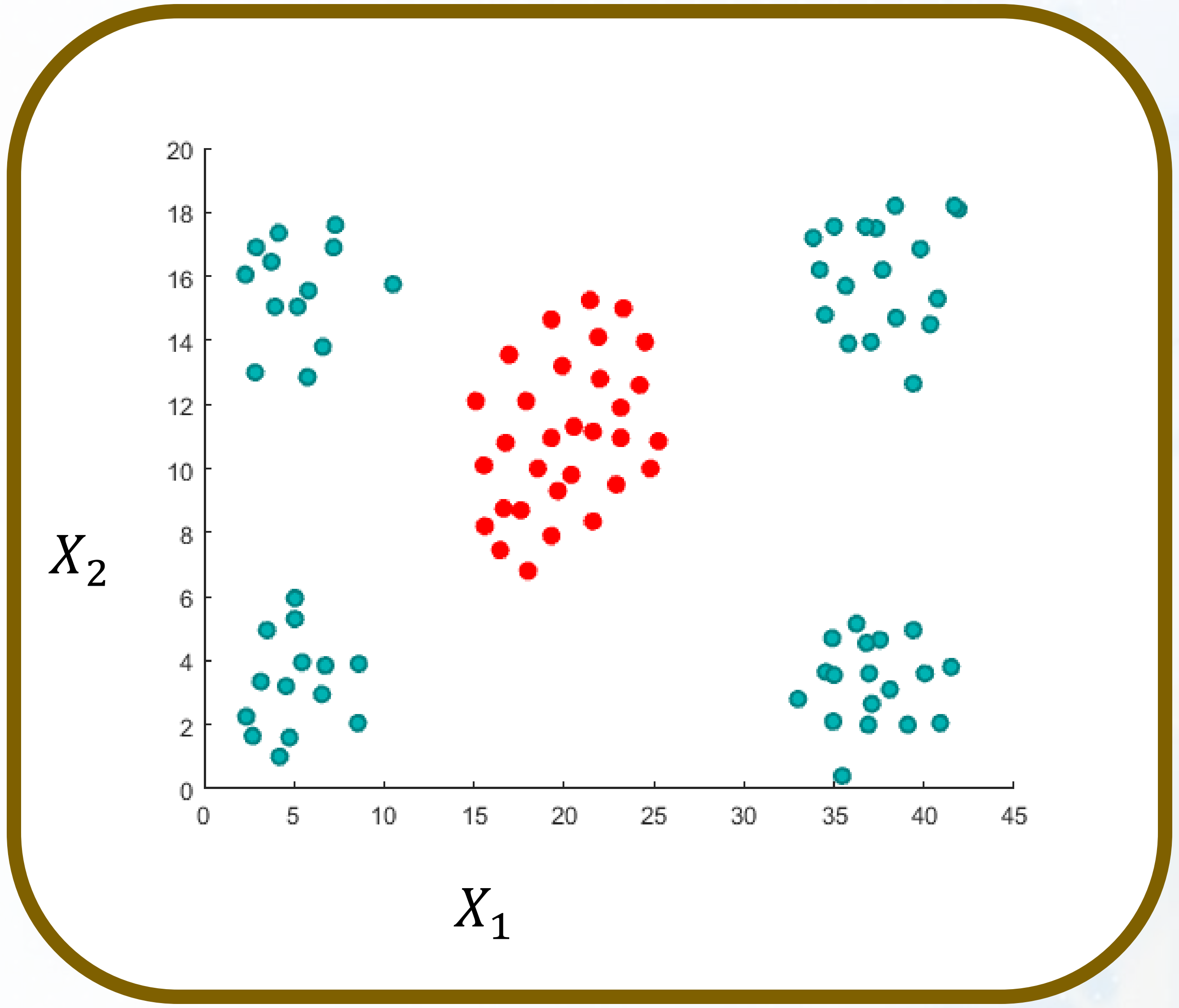




$$X \xrightarrow[(X_1^2, X_2^2)]{} Z$$

In x Space

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j x_i x_j^T$$



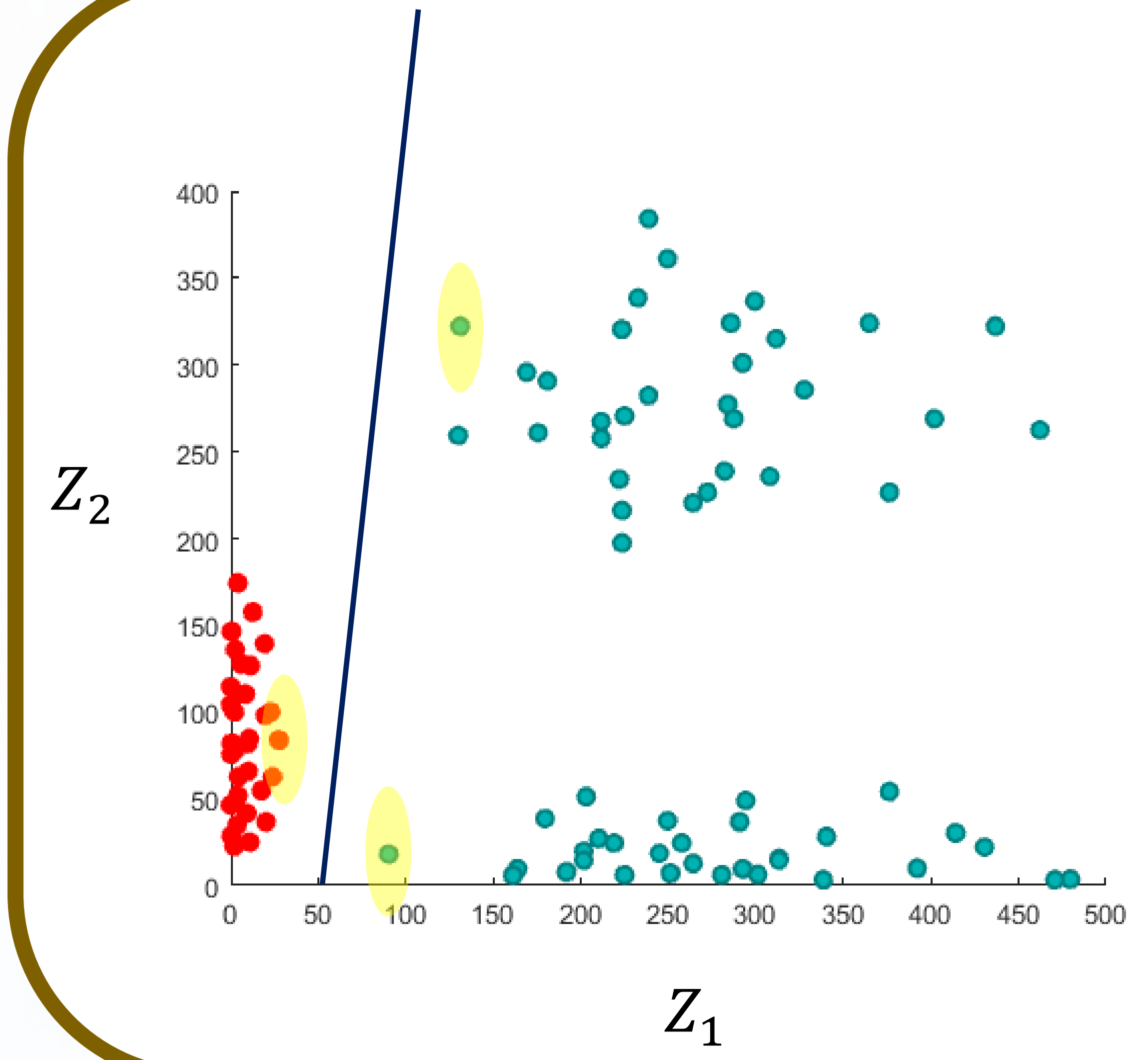
$let's\ say\ \boldsymbol{x}\ is\ n \times d$

$\boldsymbol{xx^T}$ will be $\mathrm{n} \times n$

If I add millions of dimensions to $\boldsymbol{x}$, would it affect the final size of $\boldsymbol{xx^T}$?

In z Space

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j z_i z_j^T$$



In x Space, They are Called Pre-images of Support Vectors