MACHINE LEARNING

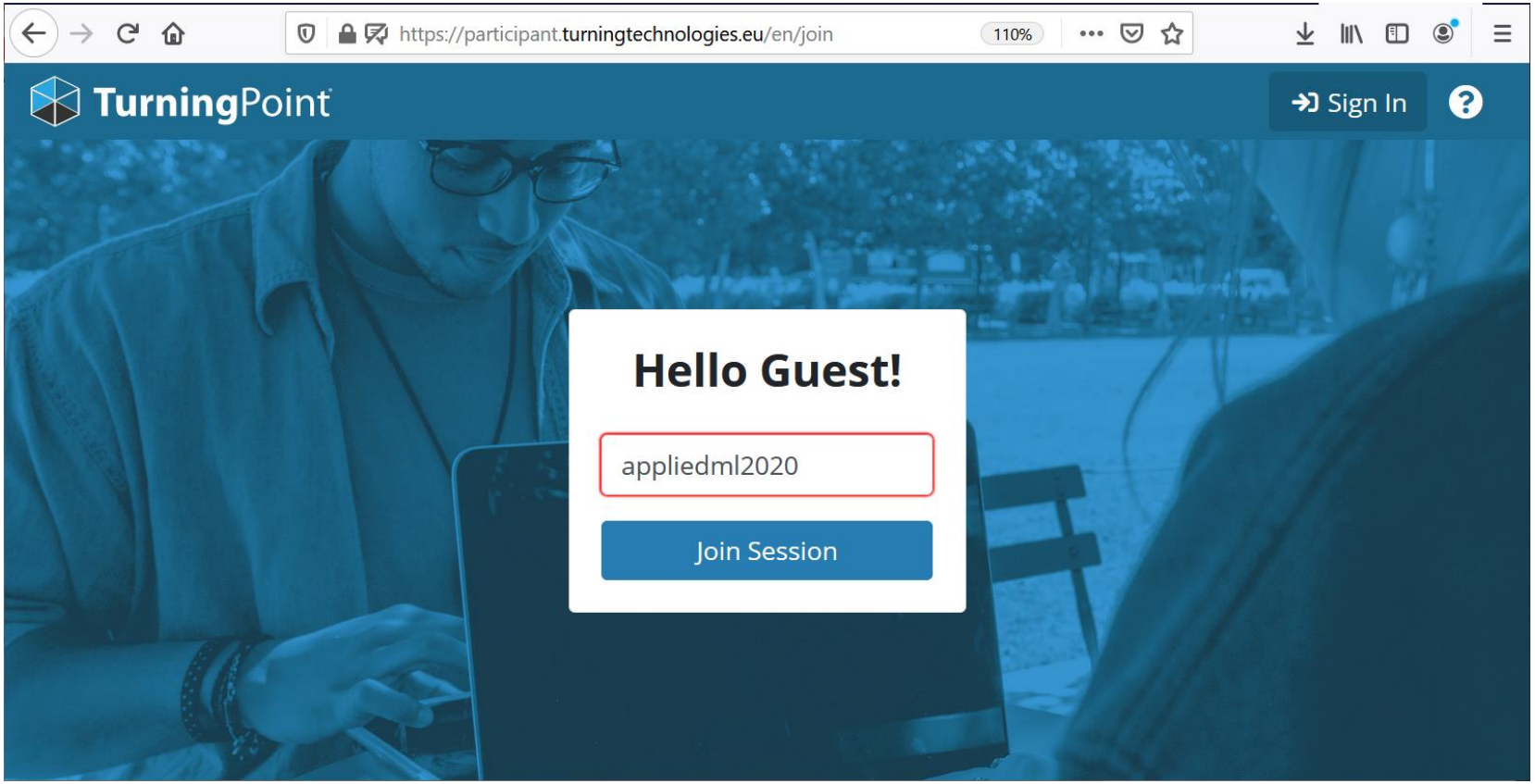
Support Vector Machine
For Classification

Interactive Lecture

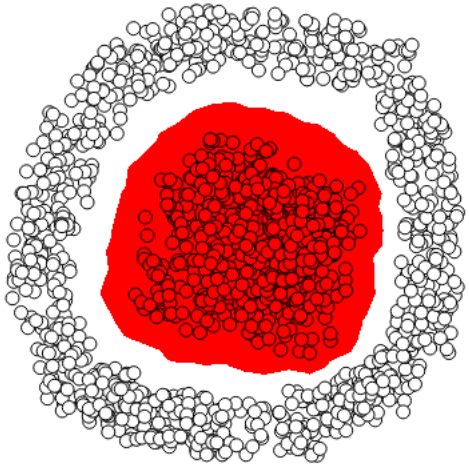
Launch polling system

https://participant.turningtechnologies.eu/en/join

Acces as GUEST and enter the session id: *appliedml2020*

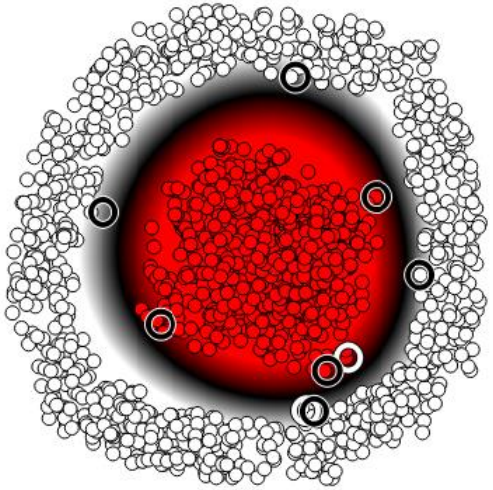


Can you recognize the classifiers?



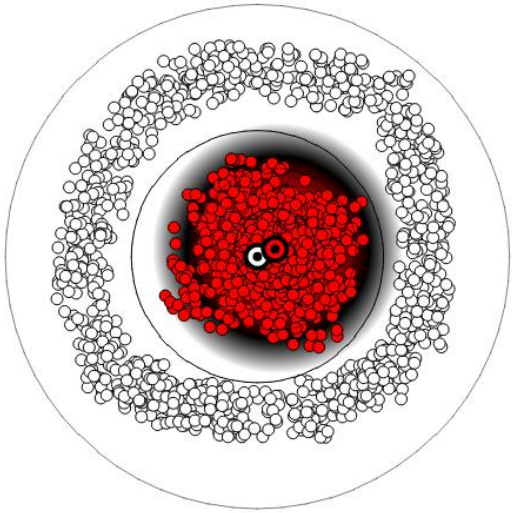
?

KNN



?

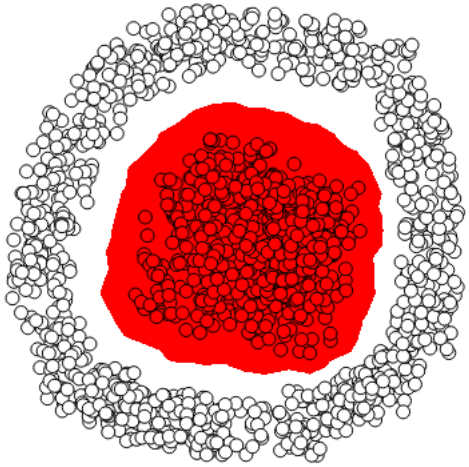
SVM



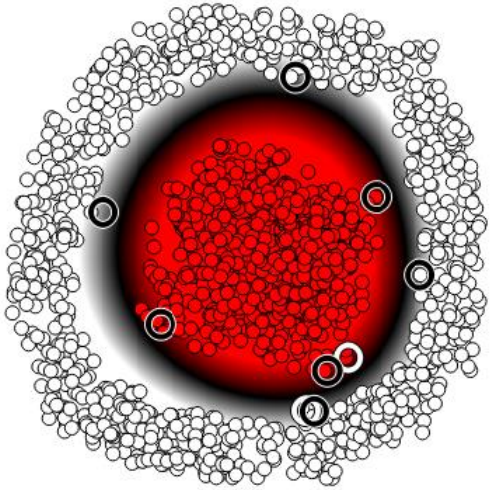
?

GMM

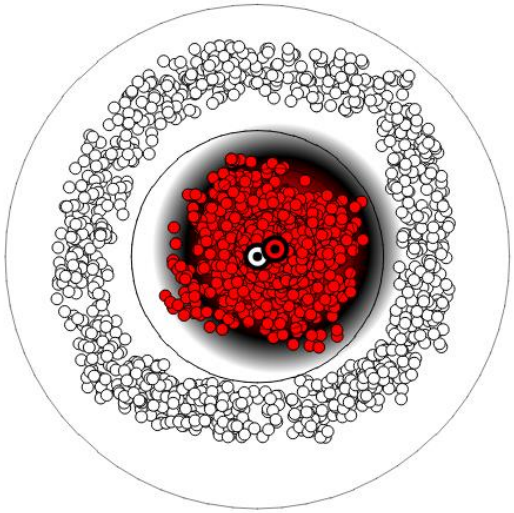
Which of the three solutions requires the least number of parameters for prediction at testing time?



KNN

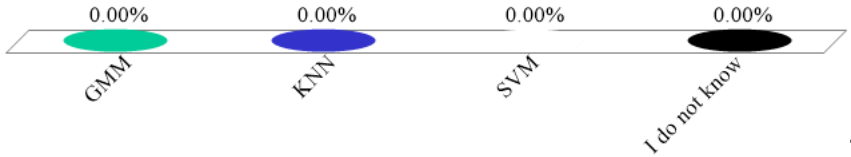


SVM

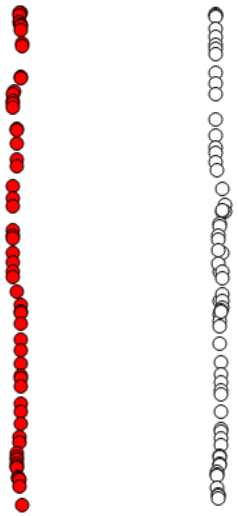


GMM

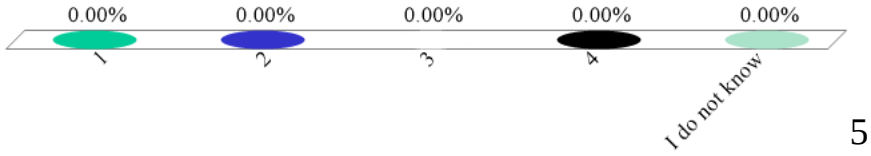
A. GMM
B. KNN
C. SVM
D. I do not know



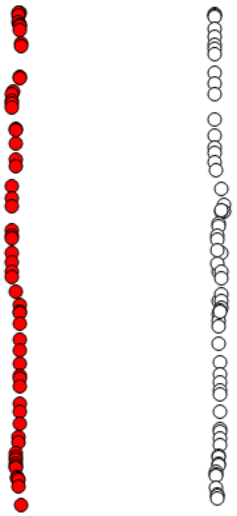
How many Support Vectors (SV) do you need at minimum when using linear SVM?



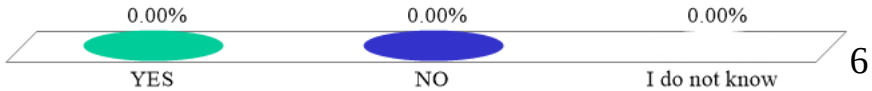
A. 1
B. 2
C. 3
D. 4
E. I do not know



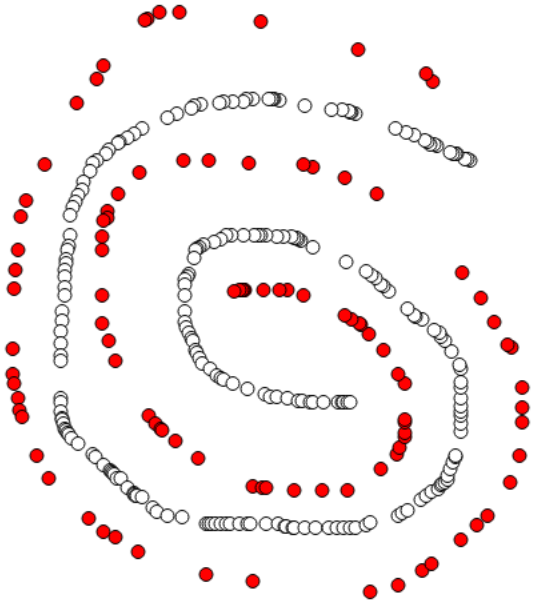
Is the solution unique?



A. YES
B. NO
C. I do not know



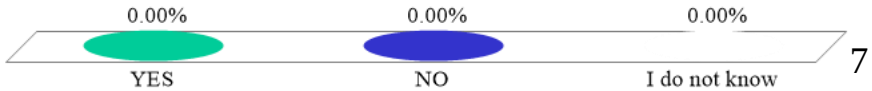
Is the solution unique?



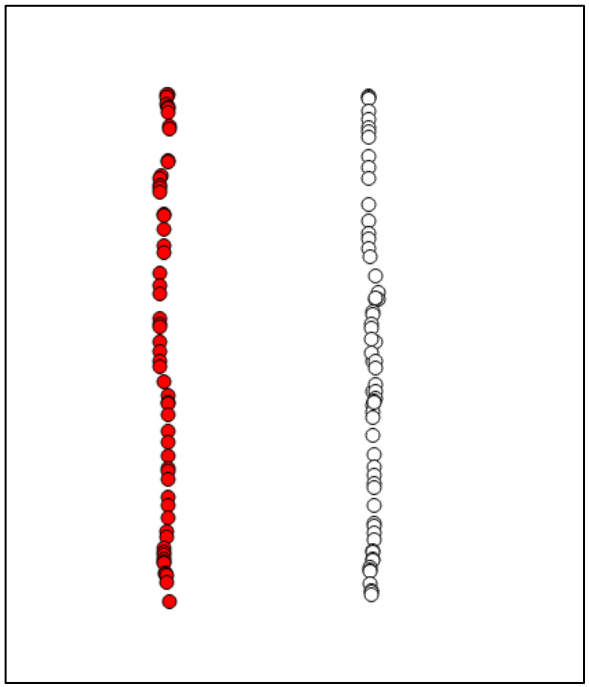
A. YES

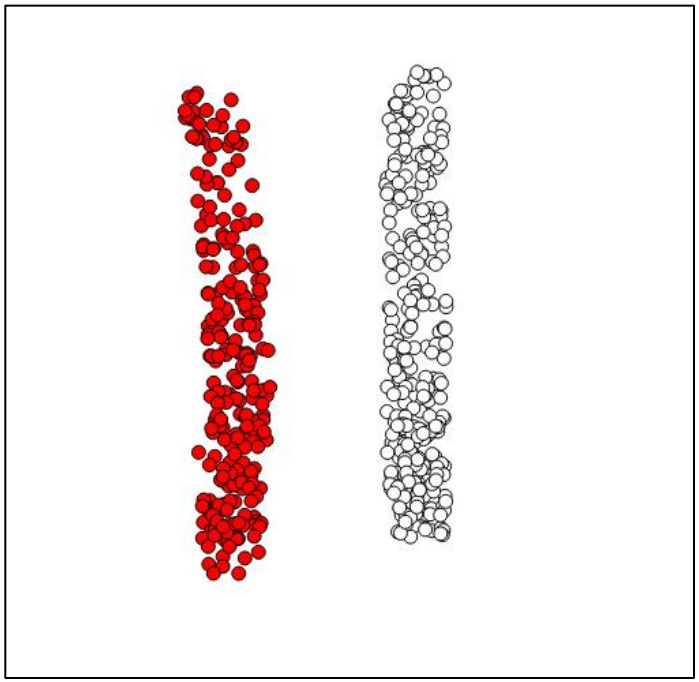
B. NO

C. I do not know



Will the number of Support Vectors (SV) change as we have more datapoints?

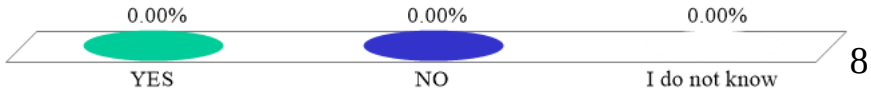




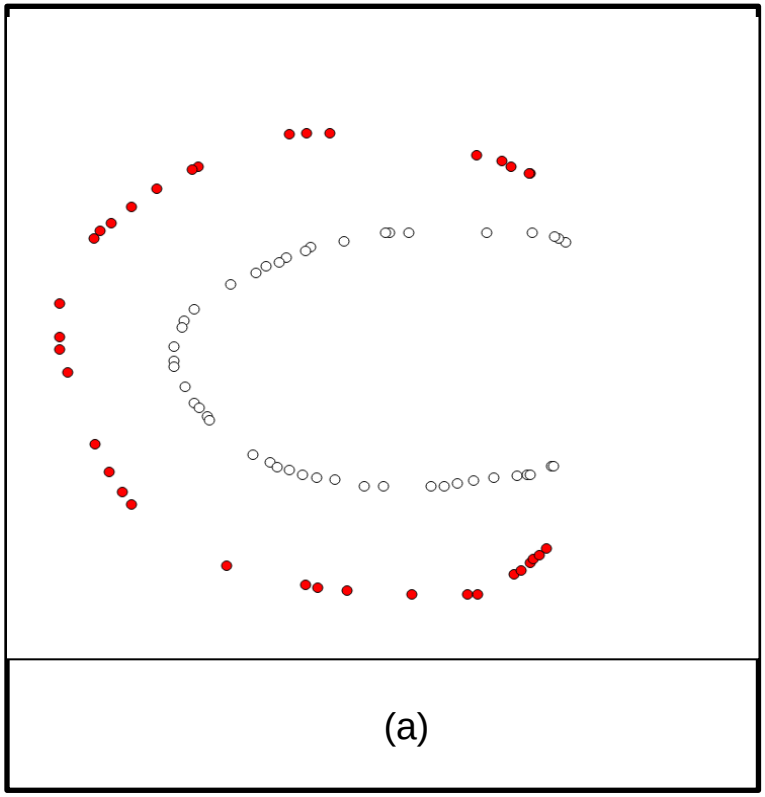
A. YES

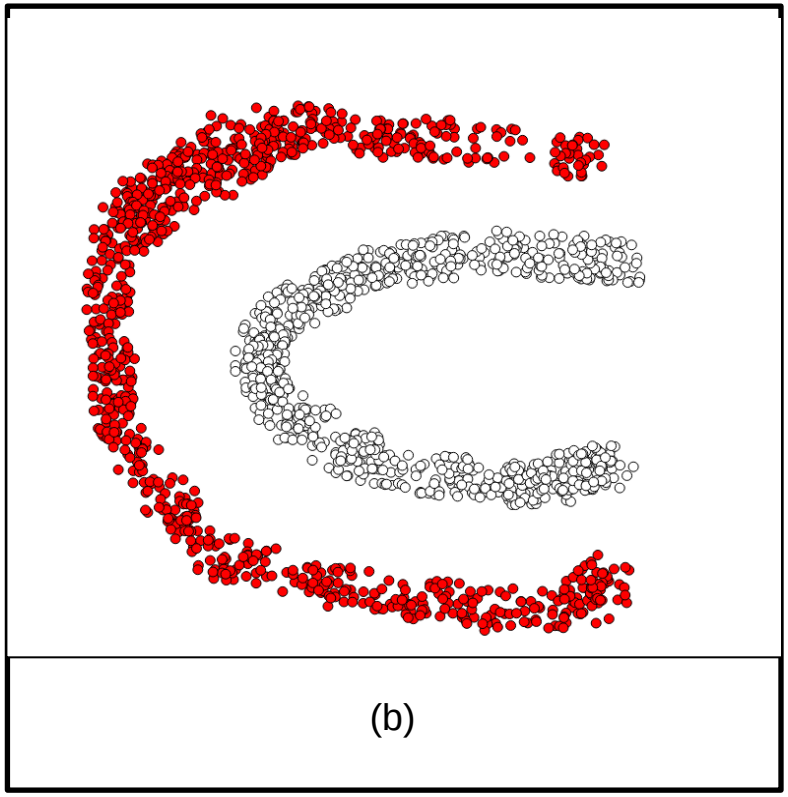
B. NO

C. I do not know

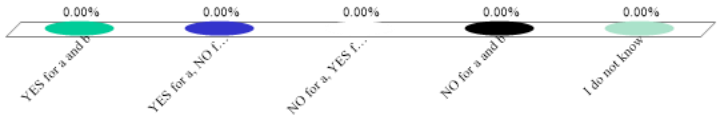


Will the number of Support Vectors (SV) change as we have more datapoints?

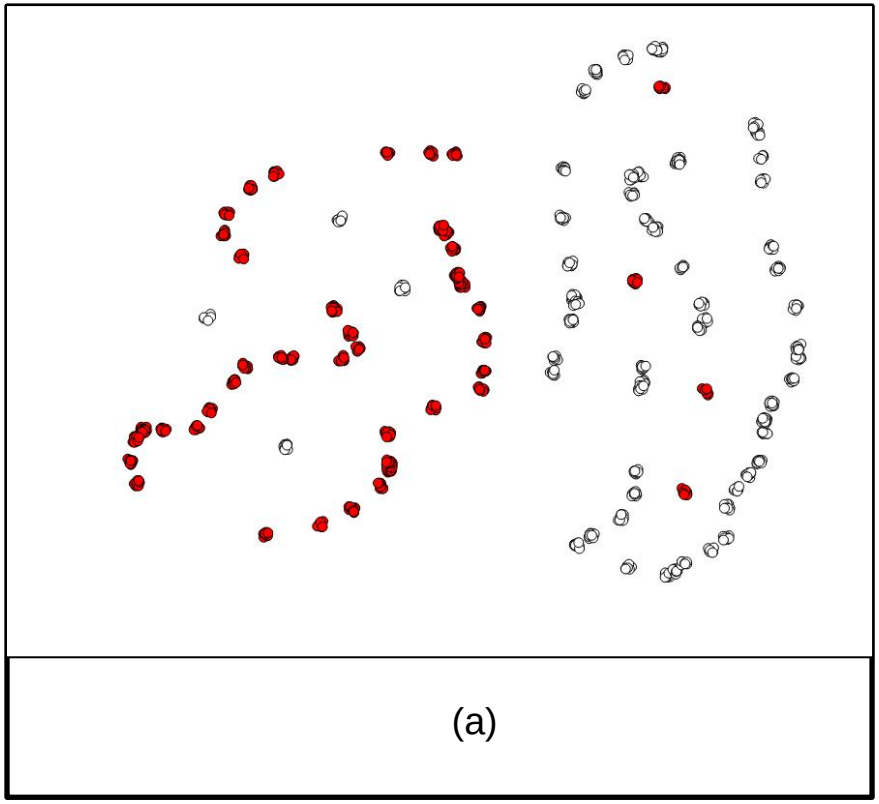


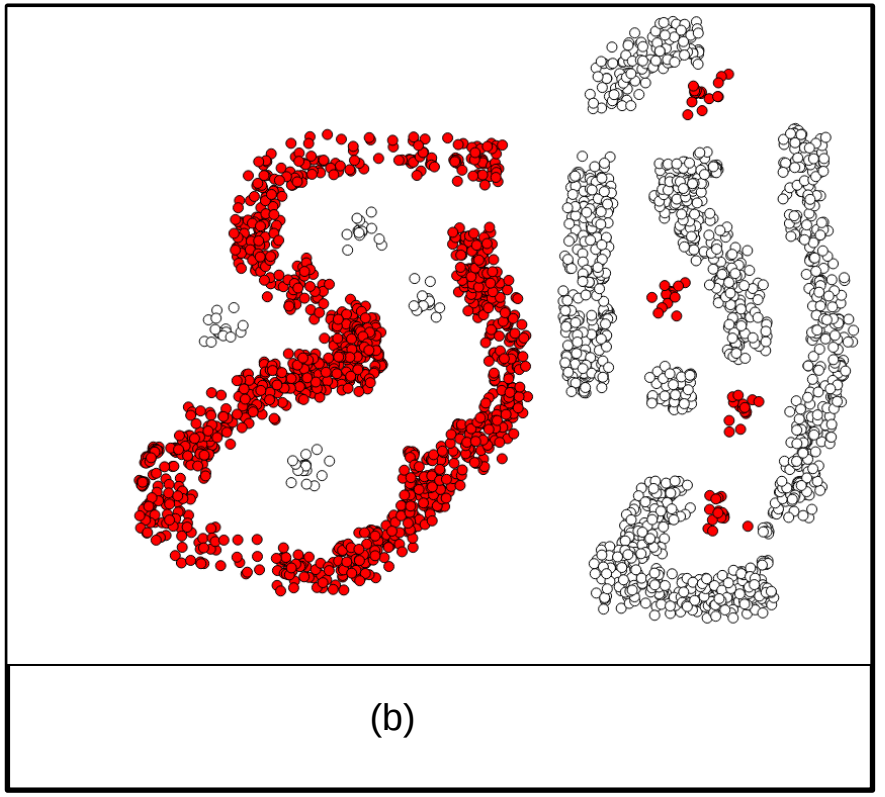


A. YES for a and b
B. YES for a, NO for b
C. NO for a, YES for b
D. NO for a and b
E. I do not know

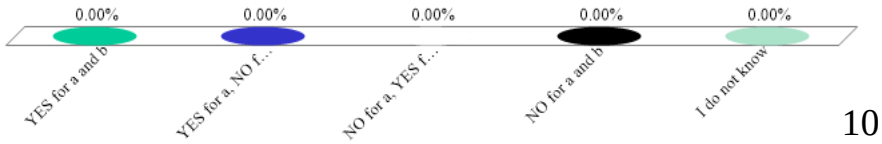


Will the number of Support Vectors (SV) change as we have more datapoints?

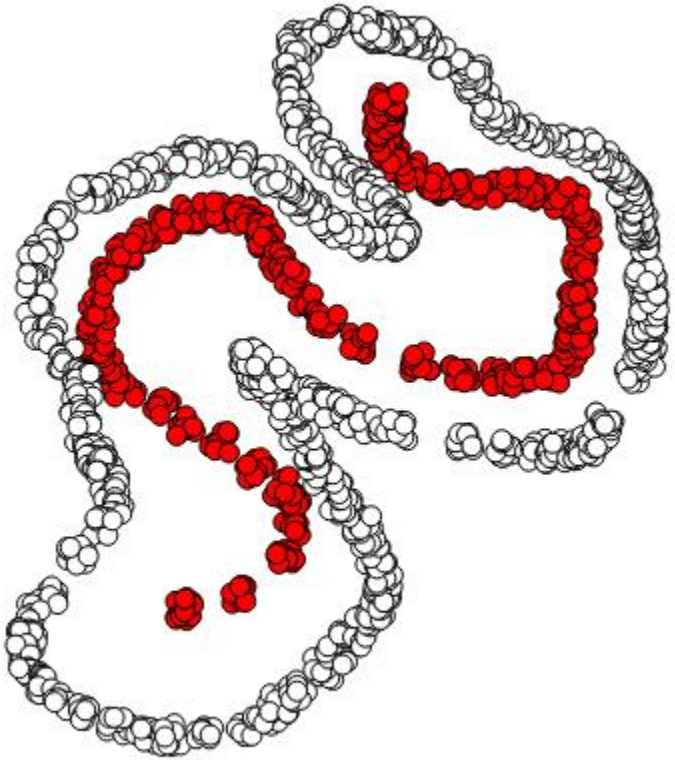




A. YES for a and b
B. YES for a, NO for b
C. NO for a, YES for b
D. NO for a and b
E. I do not know



Non-Linear Classification



The decision function:

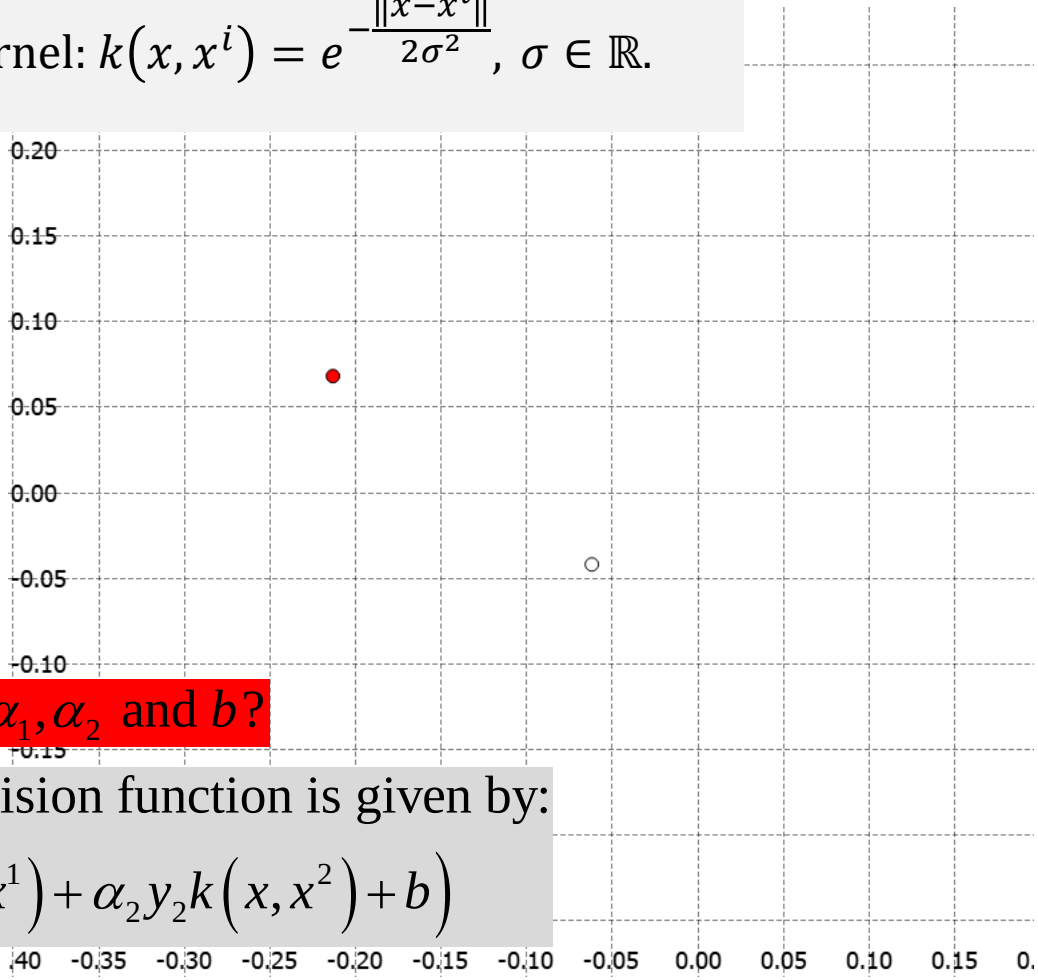
$$f(x) = sgn\left(\sum_{i=1}^{M} \alpha_i y_i k\left(x, x^i\right) + b\right)$$

How can we build the decision boundary?

The decision function is given by:

$$f(x) = sgn\left(\sum_{i=1}^{M} \alpha_i y_i k\left(x, x^i\right) + b\right)$$

RBF (Gaussian) kernel: $k\left(x, x^i\right) = e^{-\frac{\|x - x^i\|}{2\sigma^2}}$, $\sigma \in \mathbb{R}$.



What are α_1, α_2 and b?

With 2 points, the decision function is given by:

$$f(x) = sgn\left(\alpha_1 y_1 k\left(x, x^1\right) + \alpha_2 y_2 k\left(x, x^2\right) + b\right)$$

The decision function is given by:

$$f(x) = sgn\left(\alpha_1 y_1 k\left(x, x^1\right) + \alpha_2 y_2 k\left(x, x^2\right) + b\right)$$

To compute α_i use:

$$\sum_{i=1}^{M} \alpha_i y^i = 0 \;\&\; y^1 = 1, y^2 = -1 \Rightarrow \alpha_1 = \alpha_2$$

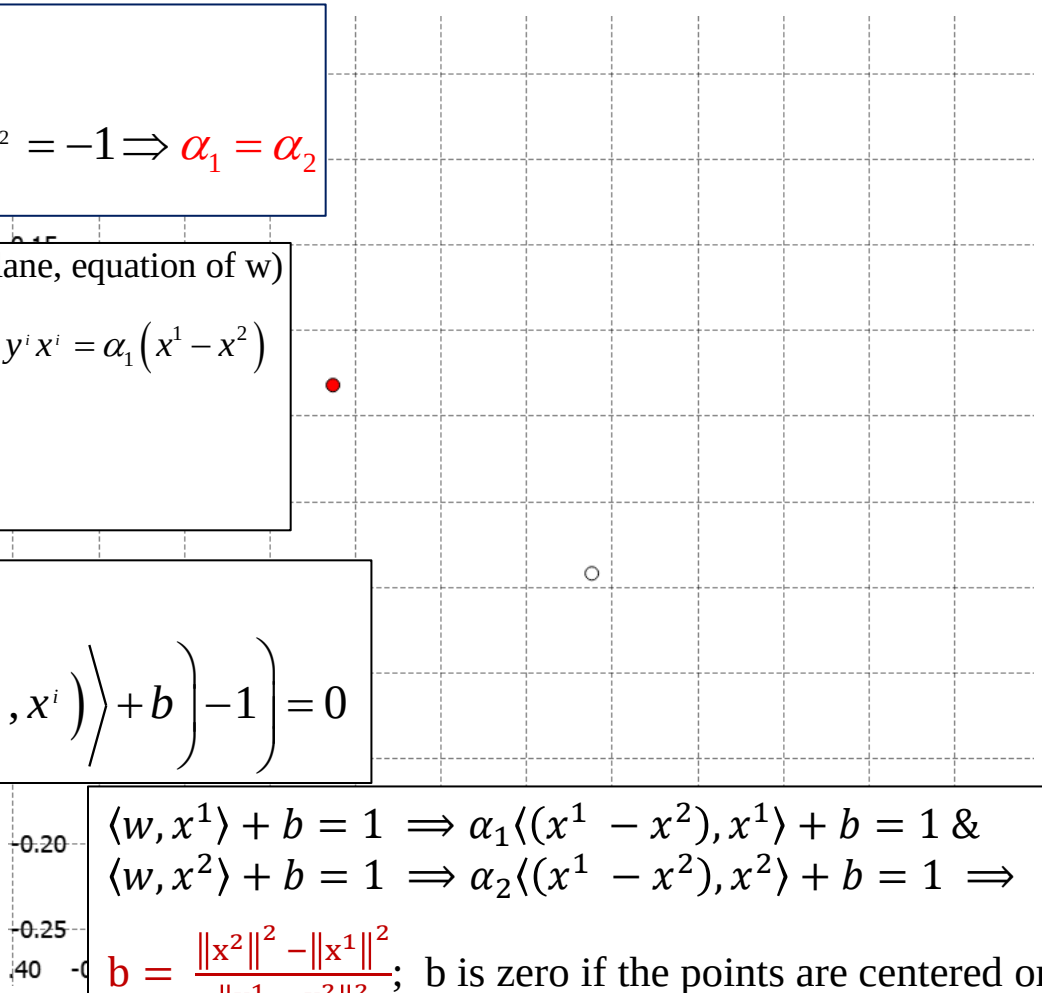
We know (distance to hyperplane, equation of w)

$$\left\|x^1 - x^2\right\| = \frac{2}{\|w\|} \quad \& \quad w = \sum_{i=1}^{M} \alpha_i y^i x^i = \alpha_1\left(x^1 - x^2\right)$$

$$\Rightarrow \alpha_1 = \alpha_2 = \frac{2}{\left\|x^1 - x^2\right\|^2}$$

To compute b

Use $\left(y^i \left(\left\langle \sum_{i=1}^{M} \alpha_i y^i k\left(x^i, x^i\right) \right\rangle + b \right) - 1 \right) = 0$



$$\langle w, x^1 \rangle + b = 1 \Rightarrow \alpha_1 \langle (x^1 - x^2), x^1 \rangle + b = 1 \;\&$$
$$\langle w, x^2 \rangle + b = 1 \Rightarrow \alpha_2 \langle (x^1 - x^2), x^2 \rangle + b = 1 \Rightarrow$$

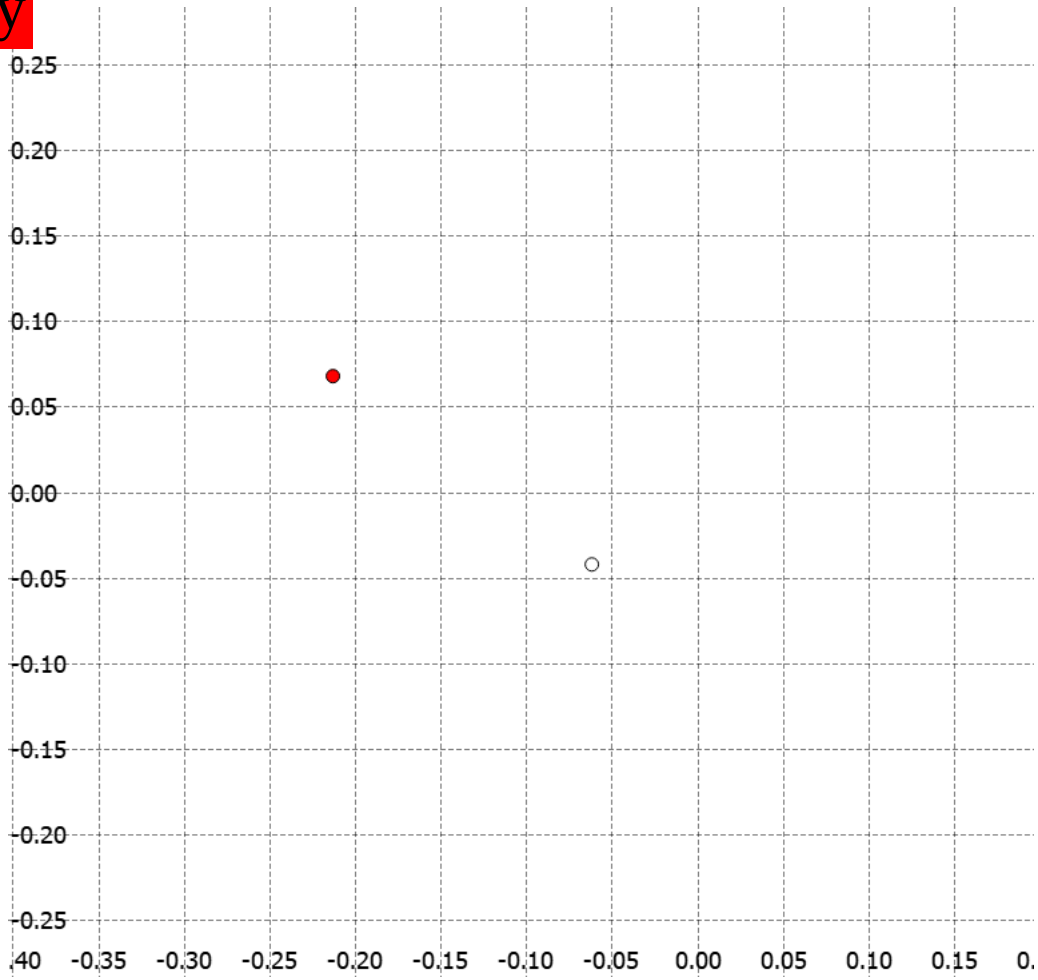
$b = \frac{\|x^2\|^2 - \|x^1\|^2}{\|x^1 - x^2\|^2}$; b is zero if the points are centered on the origin.

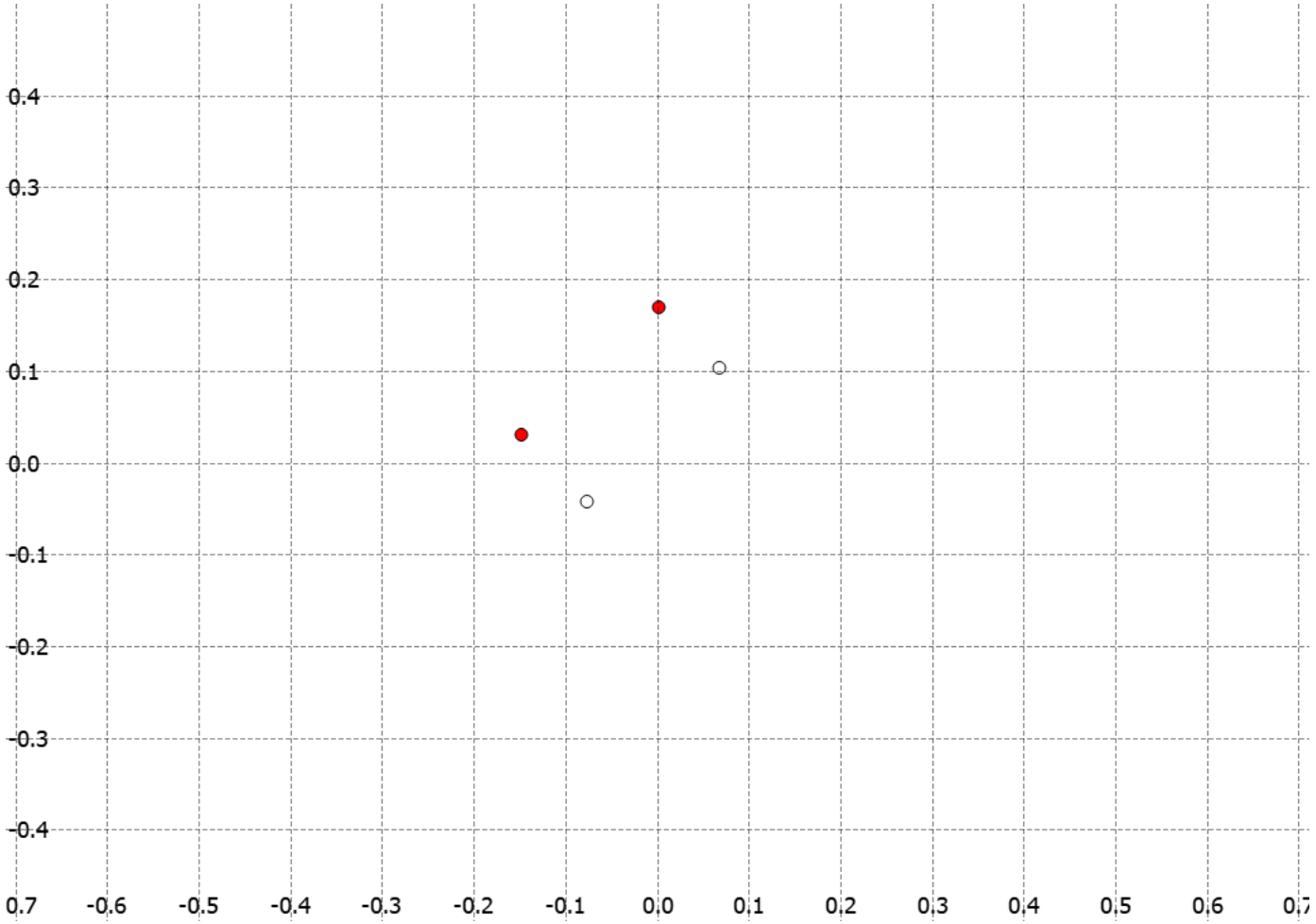
See exercise session

The decision function is given by:

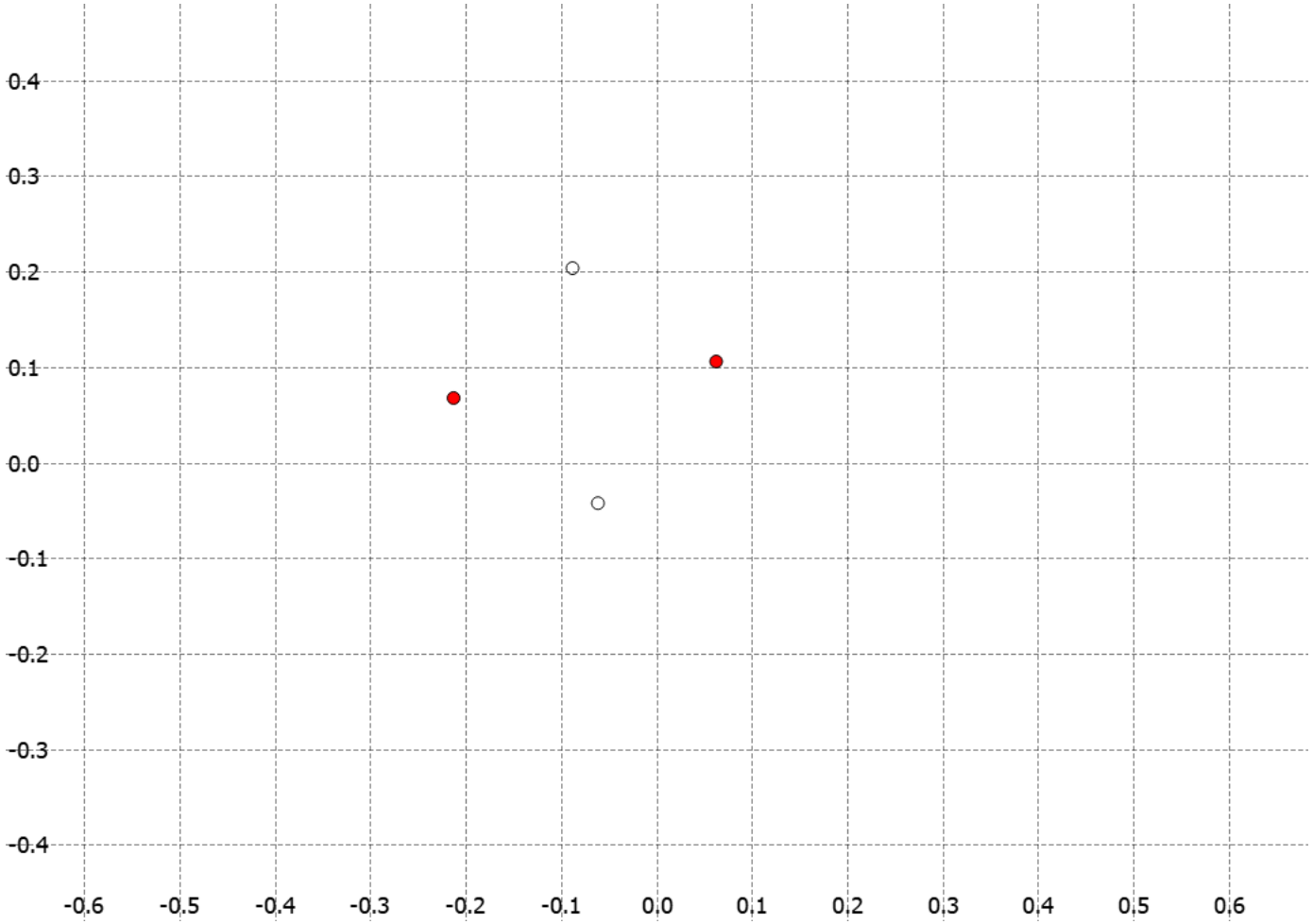
$$f(x) = sgn\left(\alpha_1 y_1 k\left(x, x^1\right) + \alpha_2 y_2 k\left(x, x^2\right) + b\right)$$

Draw the boundary





Draw the decision boundary?

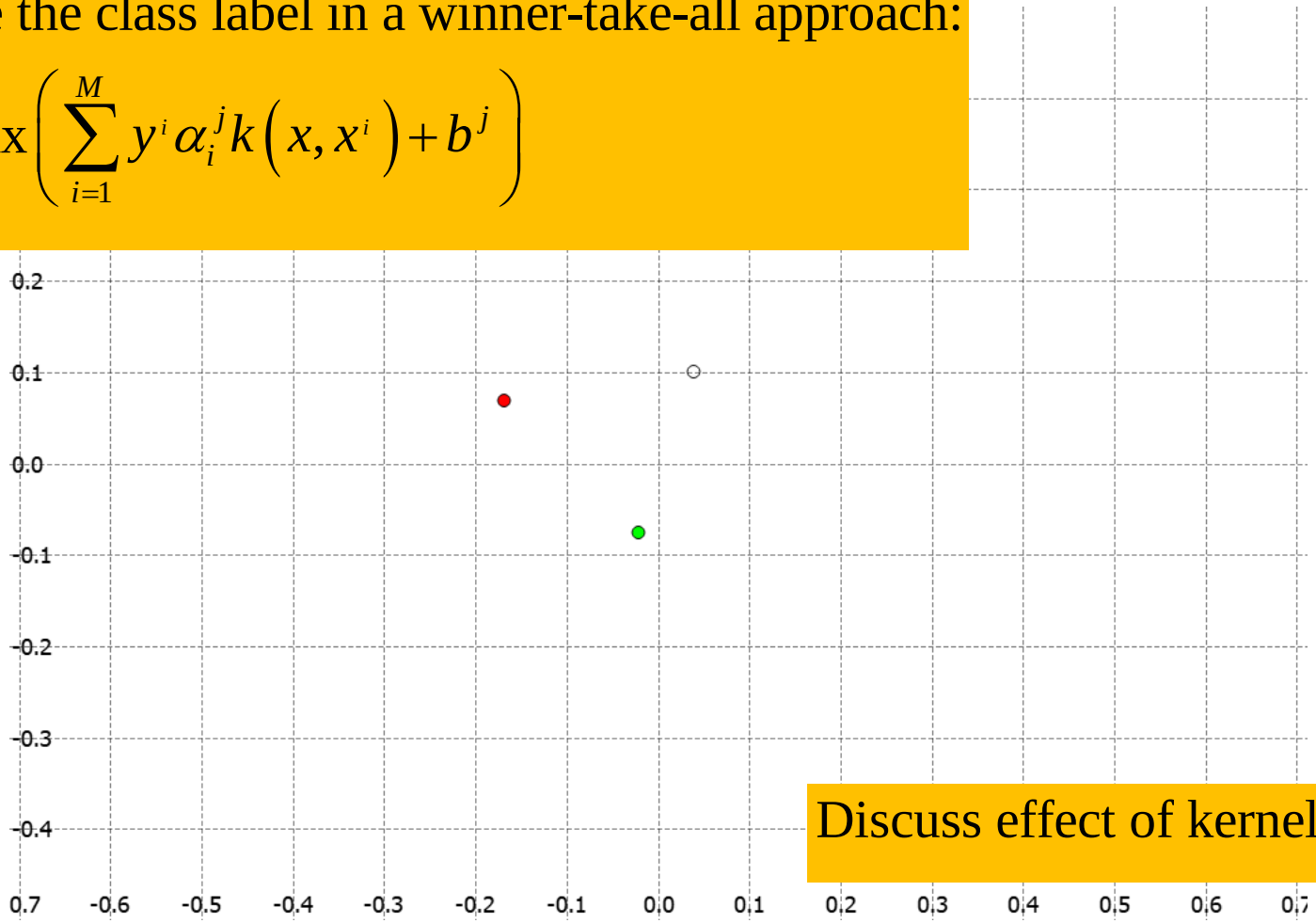


Draw the decision boundary?

Multi-Class SVM

Compute the class label in a winner-take-all approach:

$$j = \arg\max_{j=1,\ldots K} \left(\sum_{i=1}^{M} y^i \alpha_i^j k\left(x, x^i\right) + b^j \right)$$



Discuss effect of kernel width σ

Draw the decision boundary (assume RBF kernel)?

Multi-Class SVM

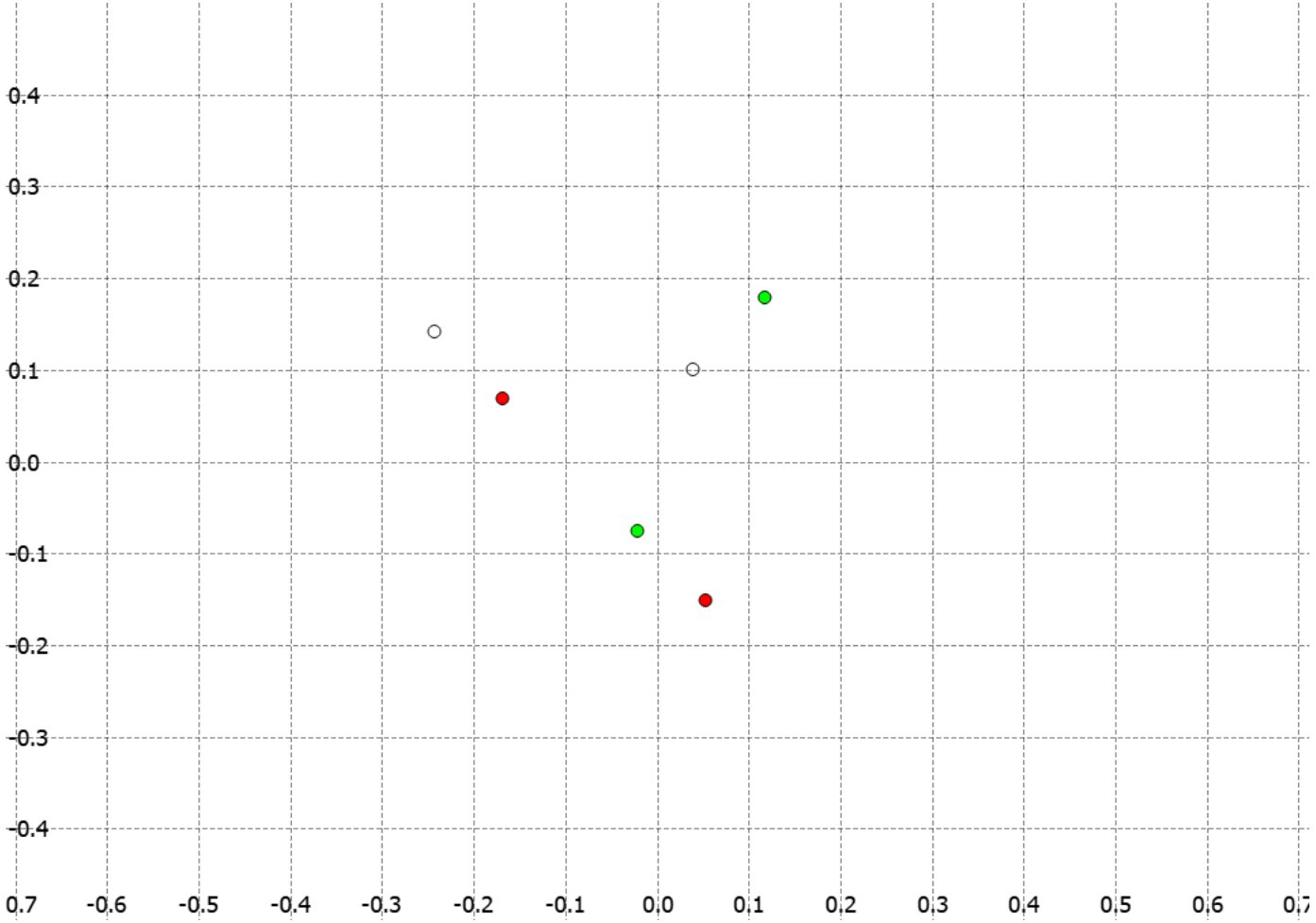
Discuss effect of kernel width σ

Compute the class label in a winner-take-all approach:

$$\text{j}=\arg\max_{j=1,\dots K}\left(\sum_{i=1}^{M} y^i \alpha_i^j \underbrace{k\left(x, x^i\right)}_{\sim 0 \text{ when } \|x-x^i\|\to\infty} + b^j\right) = \arg\max_{j=1,\dots K}\left(b^j\right)$$

The kernel width does not affect the boundary except when very far from the datapoints, when (numerically) the RBF function becomes zero and prediction is based on b

Multi-Class SVM



Draw the decision boundary (assume RBF kernel)?