Machine Learning I

MICRO-455

Classification with SVM

The Teaching Team

Fall 2024

Q1

- Computational Aspects of SVM.
  - Number of datapoints: M.
  - Data dimension: N.
  - Number of support vectors: S.

Q1 > A

- What is the number of parameters to store?

Q1 > A

- What is the number of parameters to store?

$$\boxed{f(\boldsymbol{x}) = \operatorname{sign}\left(\left(\sum_i \alpha_i y_i \langle \boldsymbol{x}, \boldsymbol{x}^i \rangle\right) + b\right)}$$

- What is the number of parameters to store?

$$\boxed{f(\boldsymbol{x}) = \operatorname{sign}\left(\left(\sum_i \alpha_i y_i \langle \boldsymbol{x}, \boldsymbol{x}^i \rangle\right) + b\right)}$$

$$\boxed{S \times N + S + 1 = S(N+1) + 1}$$

Q1 > A

- What is the number of parameters to store?

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\left(\sum_i \alpha_i y_i \langle \boldsymbol{x}, \boldsymbol{x}^i \rangle\right) + b\right)$$

$$S \times N + S + 1 = S(N+1) + 1$$

Q1 > A

- What is the number of parameters to store?

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\left(\sum_i \alpha_i y_i \langle \boldsymbol{x}, \boldsymbol{x}^i \rangle\right) + b\right)$$

$$S \times N + S + 1 = S(N+1) + 1$$

Q1 > A

- What is the number of parameters to store?

$$\boxed{f(\boldsymbol{x}) = \text{sign}\left(\left(\sum_i \alpha_i y_i \langle \boldsymbol{x}, \boldsymbol{x}^i \rangle\right) + b\right)}$$

$$\boxed{S \times N + S + 1 = S(N+1) + 1}$$

Q1 > A

- What is the number of parameters to store?

$$f(\boldsymbol{x}) = \text{sign}\left(\left(\sum_i \alpha_i y_i \langle \boldsymbol{x}, \boldsymbol{x}^i \rangle\right) + b\right)$$

$$S \times N + S + 1 = S(N+1) + 1$$

- It is possible to store one less parameter; but it is not practical.

$$\sum_i^M \alpha_i y_i = \sum_i^S \alpha_i y_i = 0$$

Q1 > B

- What is the required memory to store the the trained model?
- S=10,000; N=100; each float takes 8B.

Q1 > B

- What is the required memory to store the the trained model?
- S=10,000; N=100; each float takes 8B.

$$\boxed{S(N+1) = 10,000 \times (100+1) = 1,010,000}$$

Q1 > B

- What is the required memory to store the the trained model?
- S=10,000; N=100; each float takes 8B.

$$\boxed{S(N+1) = 10,000 \times (100+1) = 1,010,000}$$

$$\boxed{\frac{1,010,000 \times 8\text{B}}{1024 \times 1024} \approx 7.71\text{MB}}$$

Q1 > C

- Training on the dataset with M=1,000 and N=10 takes 0.1s.
- What is training time for M=1,000,000 and N=100?
- The training time complexity is assumed to be: $\mathcal{O}(MN^2)$.

Q1 > C

- Training on the dataset with M=1,000 and N=10 takes 0.1s.
- What is training time for M=1,000,000 and N=100?
- The training time complexity is assumed to be: $\mathcal{O}(MN^2)$.

$$\boxed{100,000 \times 0.1\text{s} = 10,000\text{s} = 166\frac{2}{3}\text{min} \approx 2.78\text{h} \approx 2\text{h}47\text{min}}$$

Q1 > D

- How much energy (in Wh) is required for the training phase?
- CPU consumption: 50W; Kettle: 1500W for 5min.

Q1 > D

- How much energy (in Wh) is required for the training phase?
- CPU consumption: 50W; Kettle: 1500W for 5min.

$$\boxed{2.78\text{h} \times 50\text{W} = 139\text{Wh}}$$

Q1 > D

- How much energy (in Wh) is required for the training phase?
- CPU consumption: 50W; Kettle: 1500W for 5min.

$$\boxed{2.78\mathrm{h} \times 50\mathrm{W} = 139\mathrm{Wh}}$$

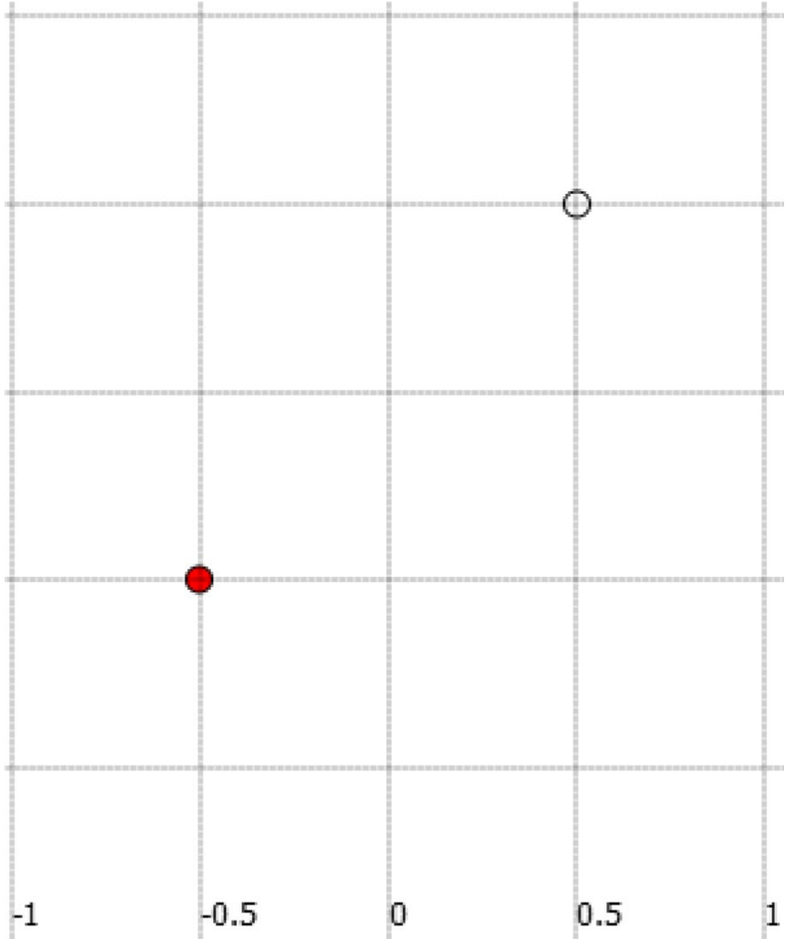
$$\boxed{\frac{5}{60}\mathrm{h} \times 1500\mathrm{W} = 125\mathrm{Wh}}$$

Q2 > A

- Compute the coefficients and the bias term of the SVM classifier for two datapoints with RBF Kernel.
- 1: X=[0.5, 0.5]; y=1
- 2: X=[-0.5, -0.5]; y=-1

$$\boxed{k(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_i\|^2}{2\sigma^2}\right)}$$

- k(X1, X2) = 0.5 = k(X2, X1)
- k(X1, X1) = 1 = k(X2, X2)



Q2 > A

- Both datapoints must be support vectors.

Q2 > A

- Both datapoints must be support vectors.

$$f(\boldsymbol{x}) = \text{sign}\left(\left(\sum_i \alpha_i y_i k(\boldsymbol{x}, \boldsymbol{x}^i)\right) + b\right) \longrightarrow \begin{cases} \sum_{i=1}^{M} \alpha_i y_i k(\boldsymbol{x}^1, \boldsymbol{x}^i) + b = 1 \\ \sum_{i=1}^{M} \alpha_i y_i k(\boldsymbol{x}^2, \boldsymbol{x}^i) + b = -1 \end{cases}$$

Q2 > A

- Both datapoints must be support vectors.

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\left(\sum_i \alpha_i y_i k(\boldsymbol{x}, \boldsymbol{x}^i)\right) + b\right) \longrightarrow \begin{cases} \sum_{i=1}^{M} \alpha_i y_i k(\boldsymbol{x}^1, \boldsymbol{x}^i) + b = 1 \\ \sum_{i=1}^{M} \alpha_i y_i k(\boldsymbol{x}^2, \boldsymbol{x}^i) + b = -1 \end{cases}$$

$$\sum_i \alpha_i y_i = 0 \longrightarrow \alpha_1 = \alpha_2$$

Q2 > A

- Both datapoints must be support vectors.

$$\boxed{f(\boldsymbol{x}) = \operatorname{sign}\left(\left(\sum_i \alpha_i y_i k(\boldsymbol{x}, \boldsymbol{x}^i)\right) + b\right) \longrightarrow \begin{cases} \sum_{i=1}^{M} \alpha_i y_i k(\boldsymbol{x}^1, \boldsymbol{x}^i) + b = 1 \\ \sum_{i=1}^{M} \alpha_i y_i k(\boldsymbol{x}^2, \boldsymbol{x}^i) + b = -1 \end{cases}}$$

$$\boxed{\sum_i \alpha_i y_i = 0 \longrightarrow \alpha_1 = \alpha_2}$$

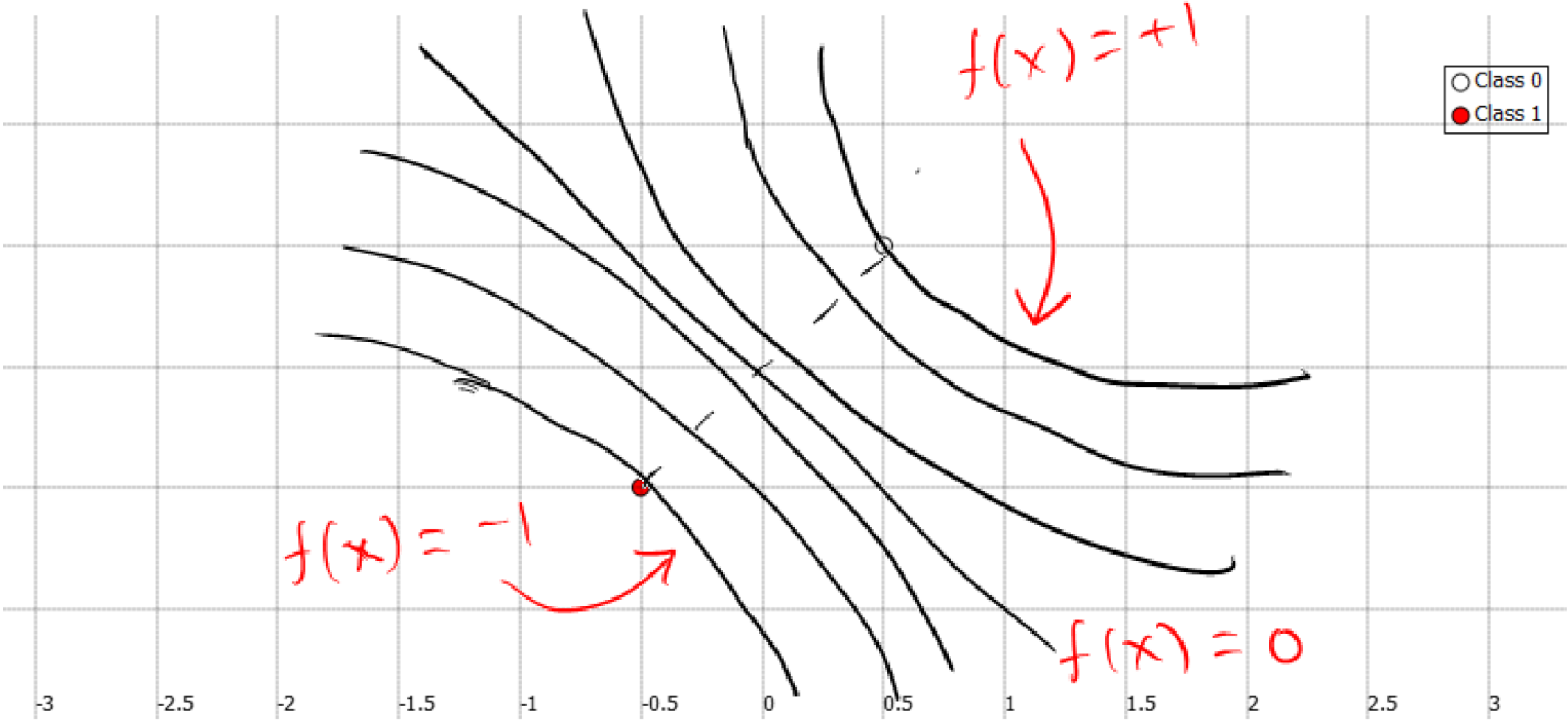
$$\boxed{\alpha_1 = \alpha_2 = 2; \quad b = 0}$$

Q2 > A

- Draw the isolines and the separating hyperplane.

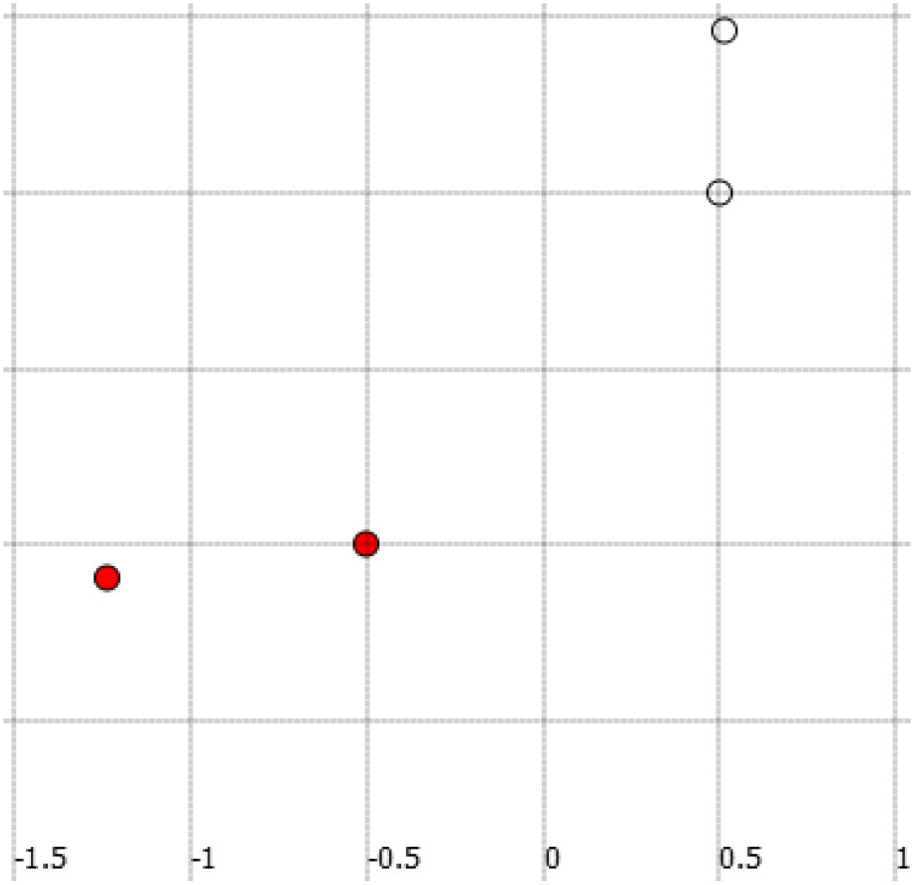
Q2 > A

- Draw the isolines and the separating hyperplane.



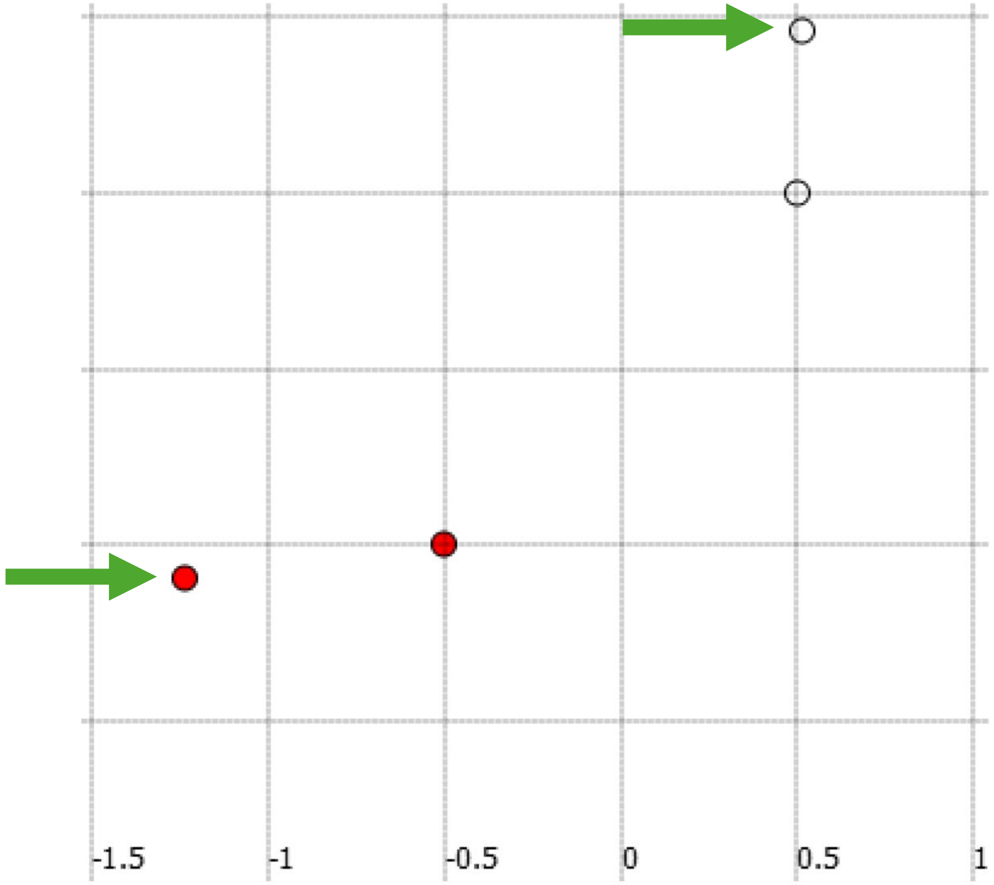
Q2 > B – Case 1

- Draw the new separating hyperplane.



Q2 > B – Case 1

- Draw the new separating hyperplane.

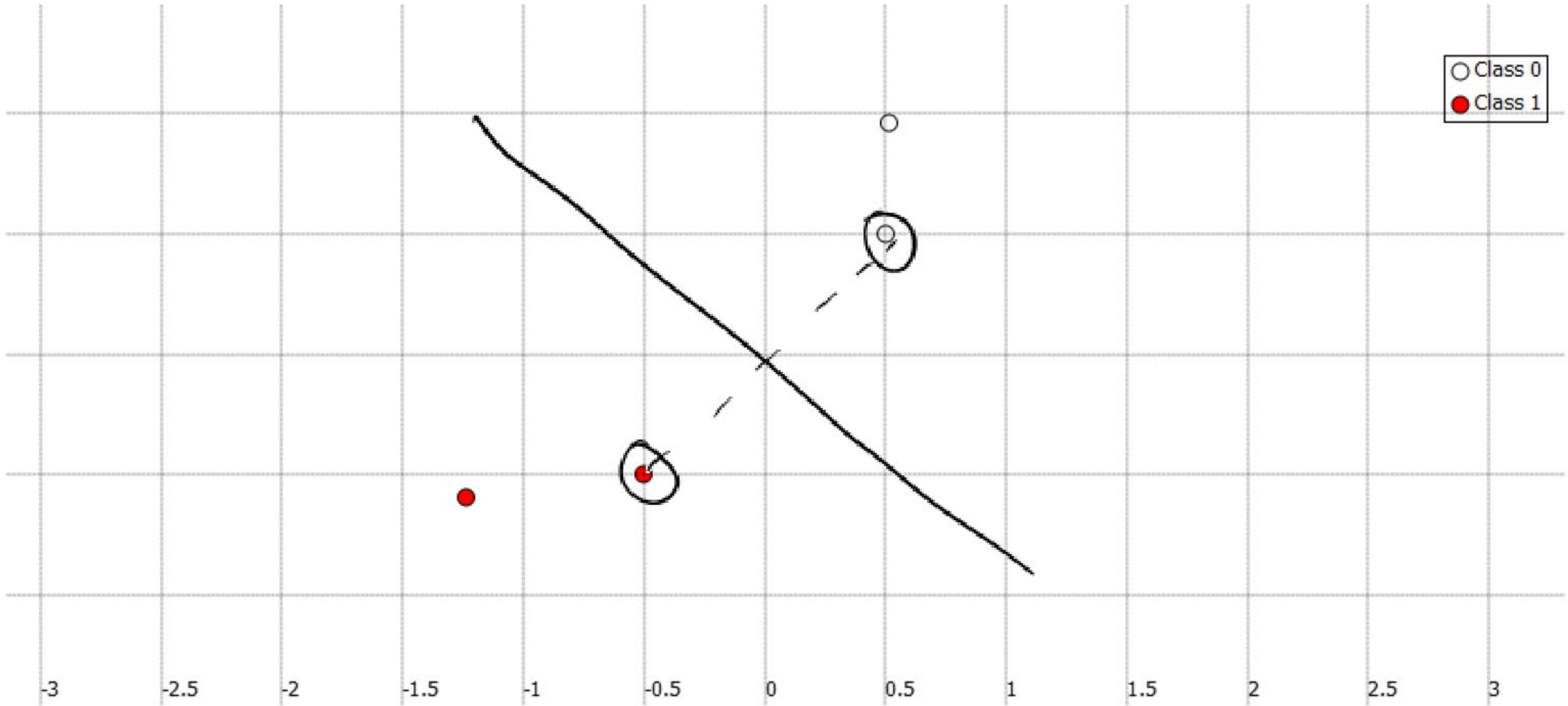


Q2 > B – Case 1

- The separating hyperplane and the support vectors not change.

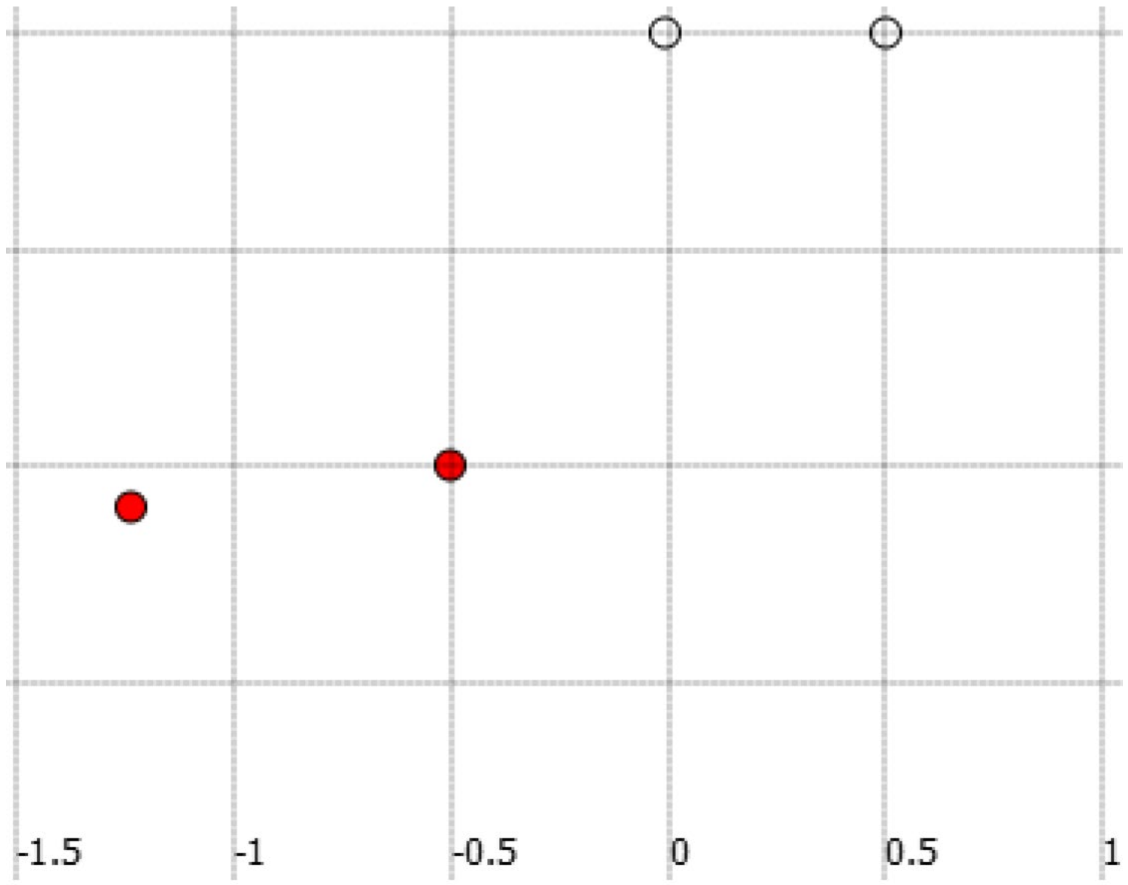
Q2 > B – Case 1

- The separating hyperplane and the support vectors not change.



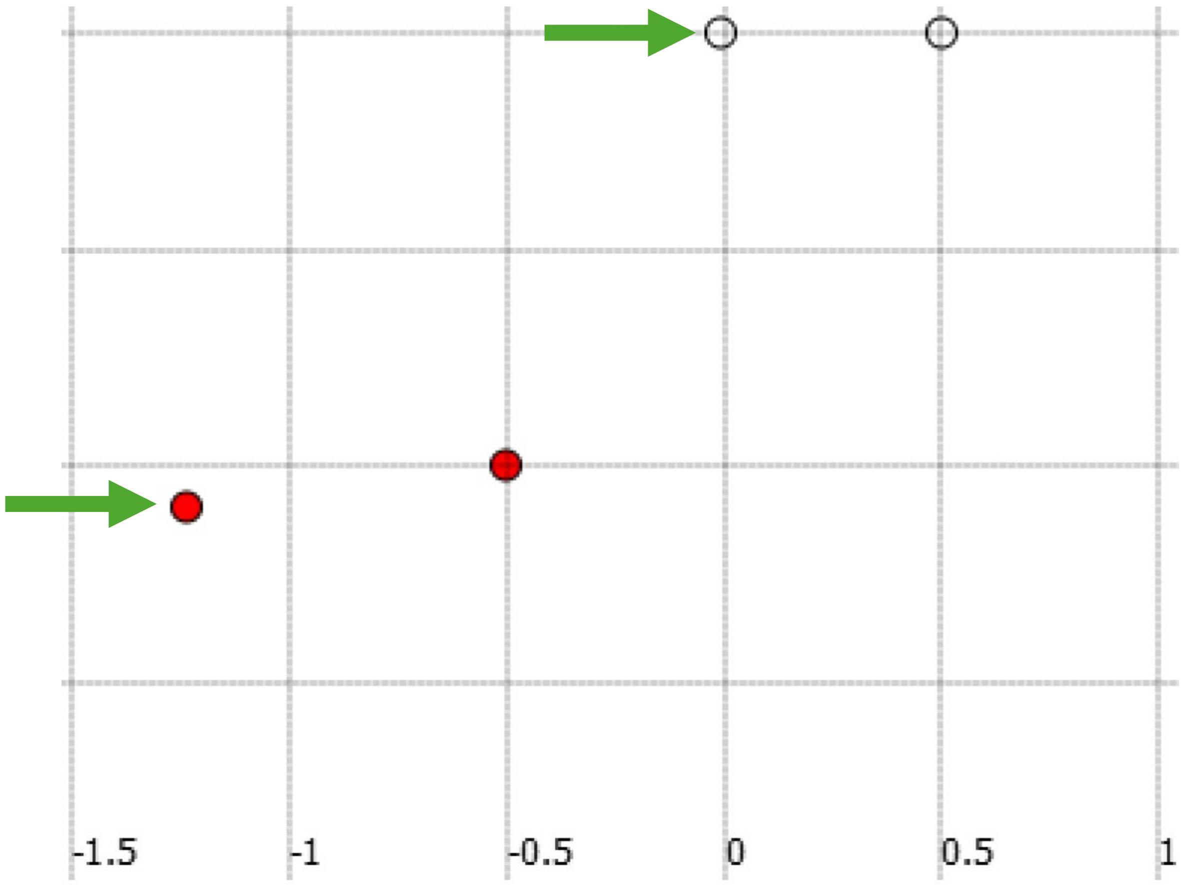
Q2 > B – Case 2

- Draw the new separating hyperplane.



Q2 > B – Case 2

- Draw the new separating hyperplane.

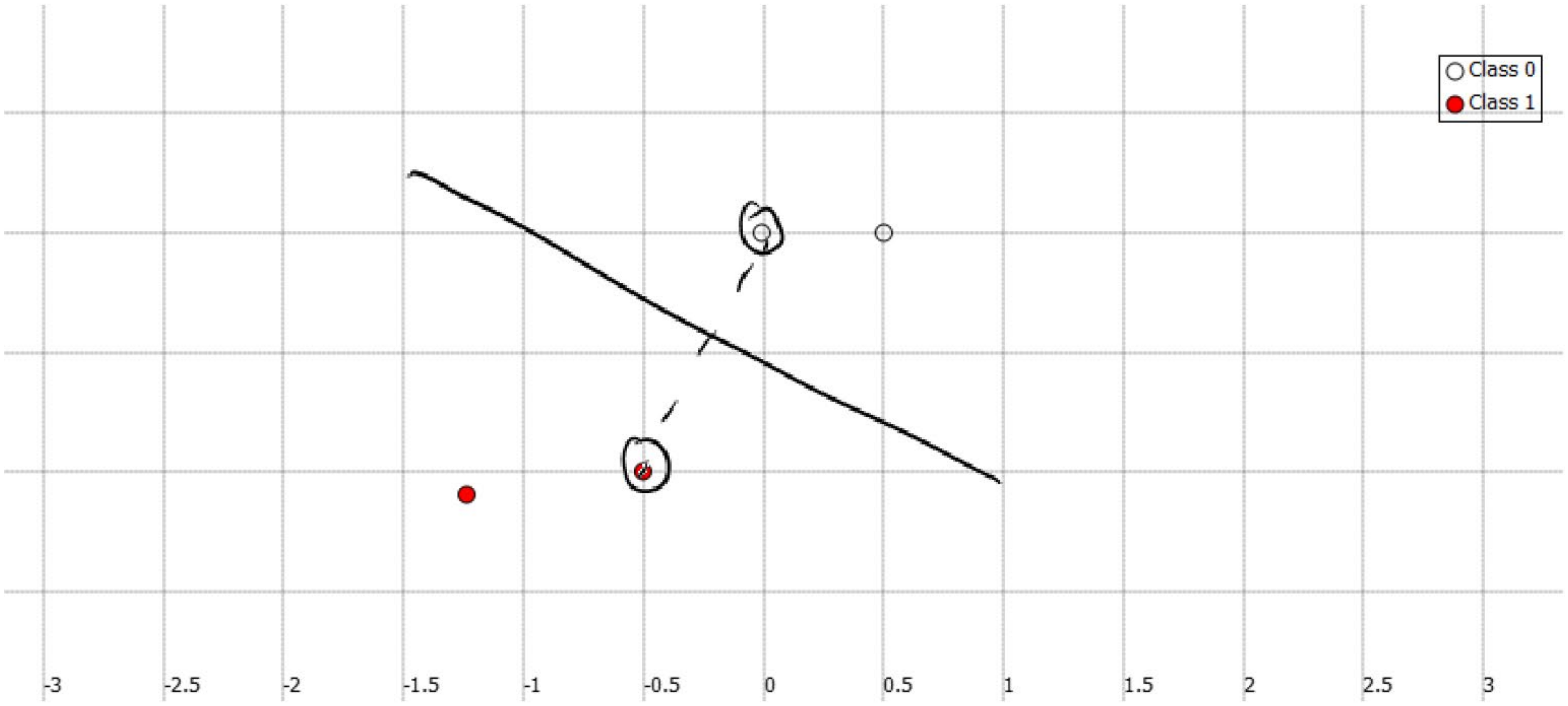


Q2 > B – Case 2

- Since the point added to the white class is inside the original margin, it now becomes a support vector instead the original point in the white class.

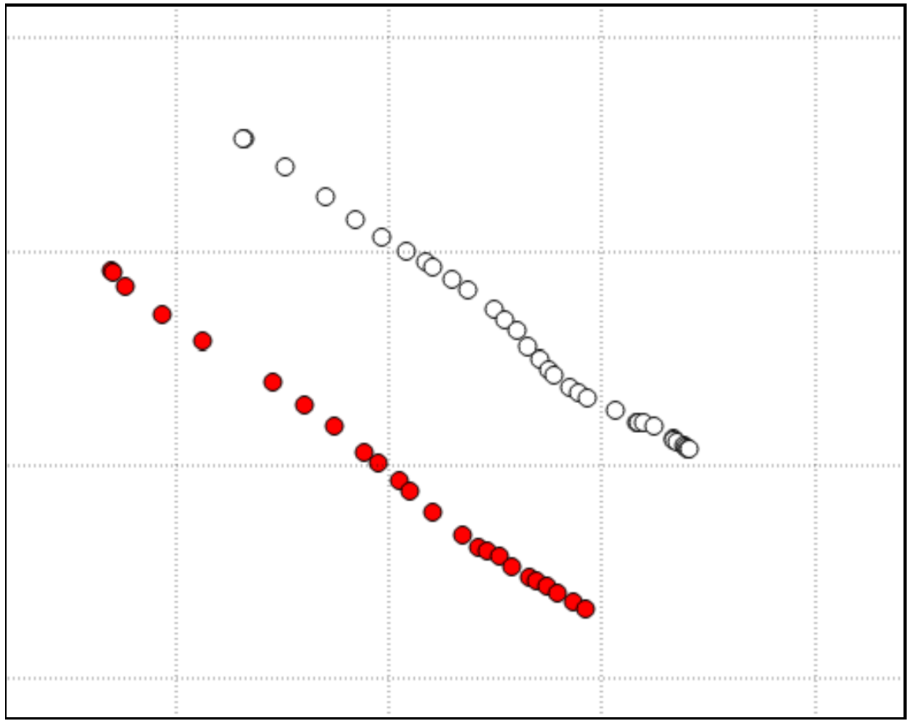
Q2 > B – Case 2

- Since the point added to the white class is inside the original margin, it now becomes a support vector instead the original point in the white class.

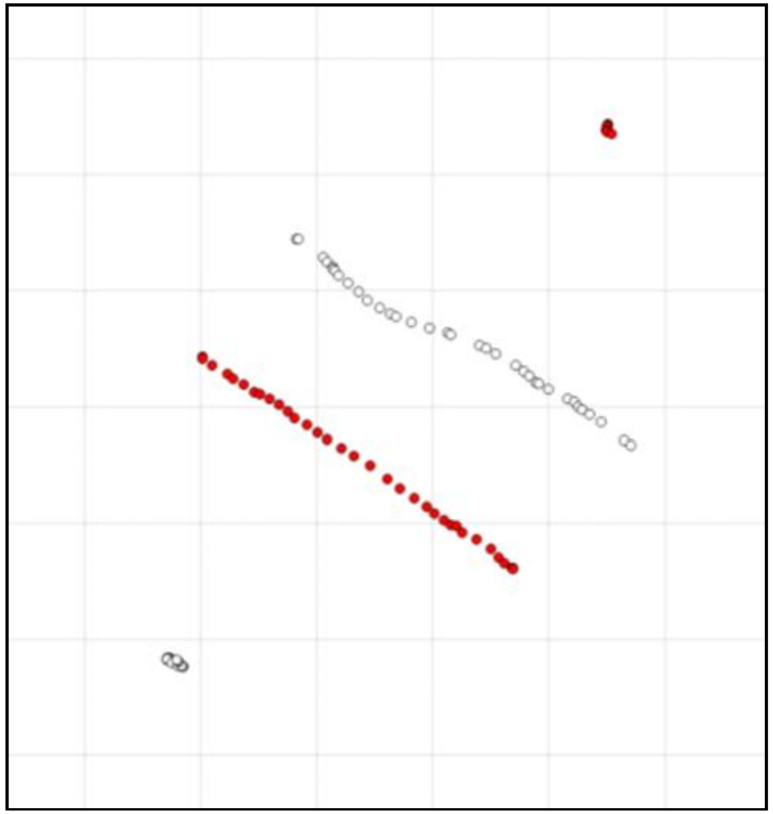


Q2 > C

- Draw the separating hyperplane and discuss the associated effects of the values of the hyperparameters (penalty factor and kernel width).



(i)



(ii)

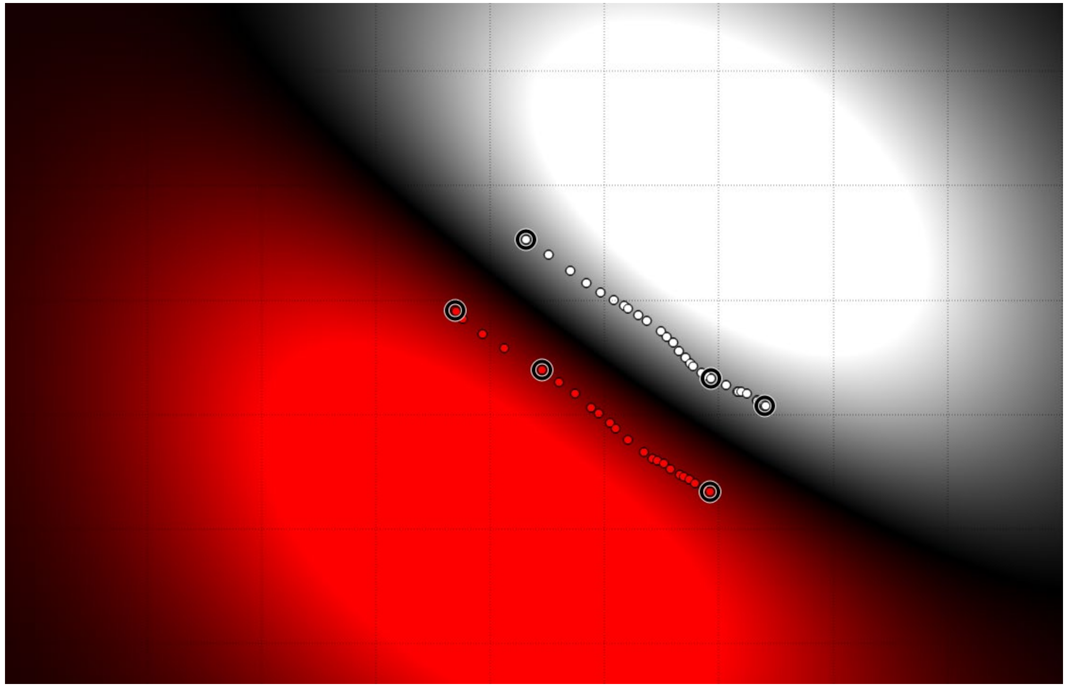
Q2 > C – Case i

- The separating line is unaffected by the value of the penalty C.
- The kernel width affects the number of support vectors.

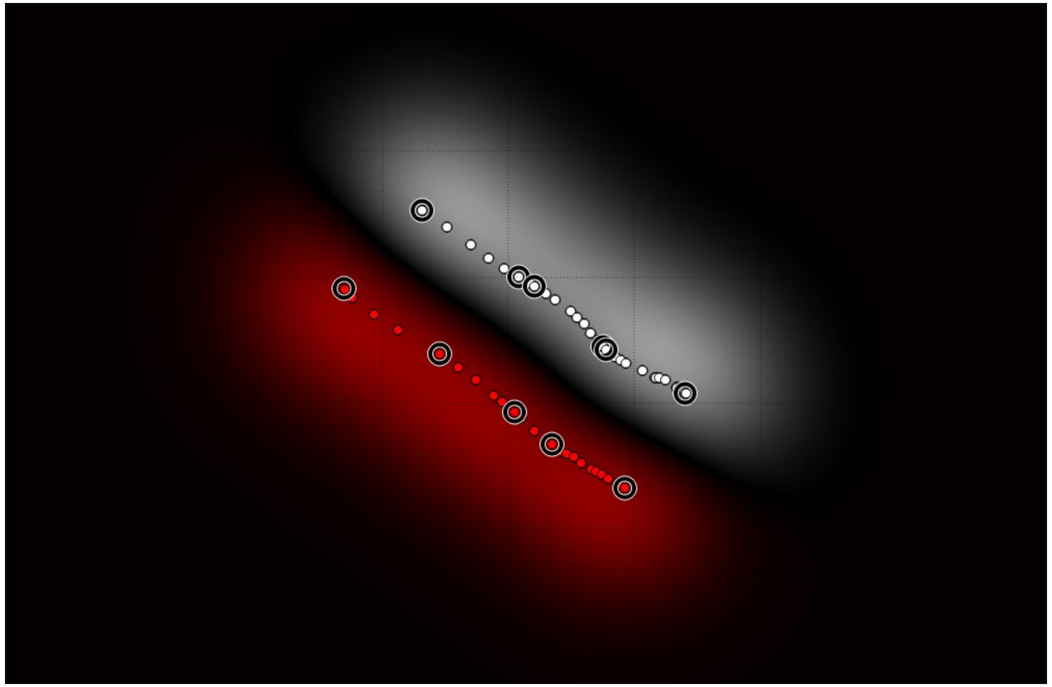
Q2 > C – Case i

- The separating line is unaffected by the value of the penalty C.
- The kernel width affects the number of support vectors.

$\sigma = 0.1$



$\sigma = 0.01$



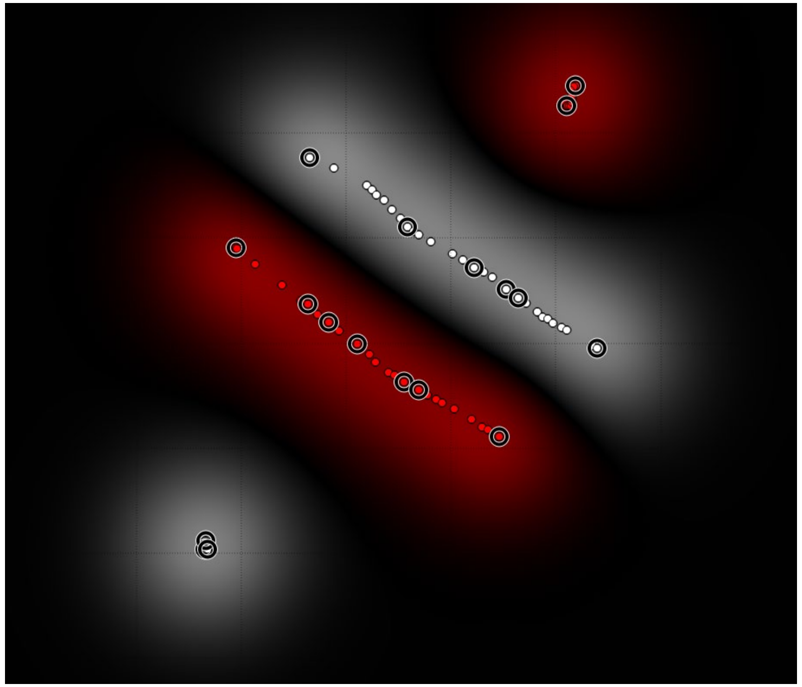
Q2 > C – Case ii

- The separating line will be influenced by both penalty C and the kernel width.

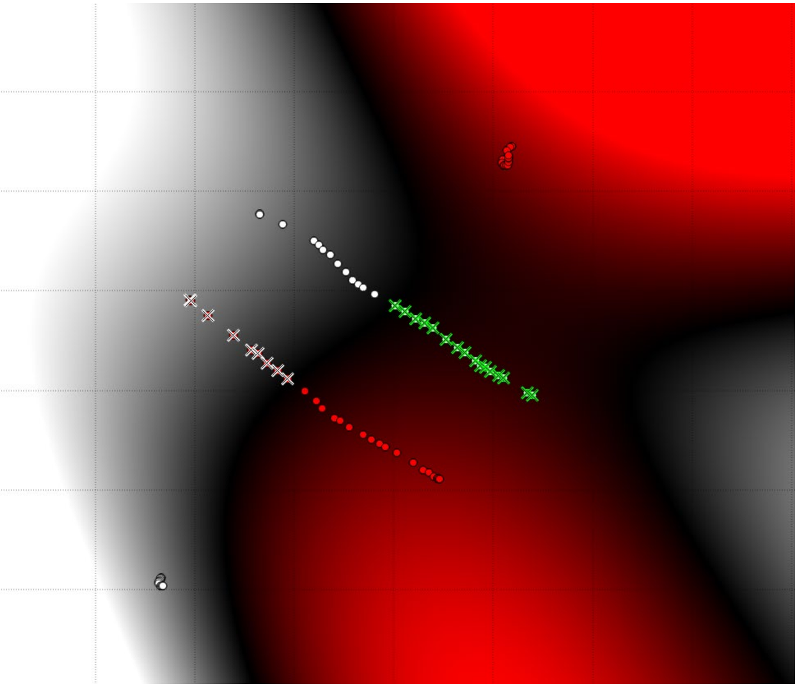
Q2 > C – Case ii

- The separating line will be influenced by both penalty C and the kernel width.

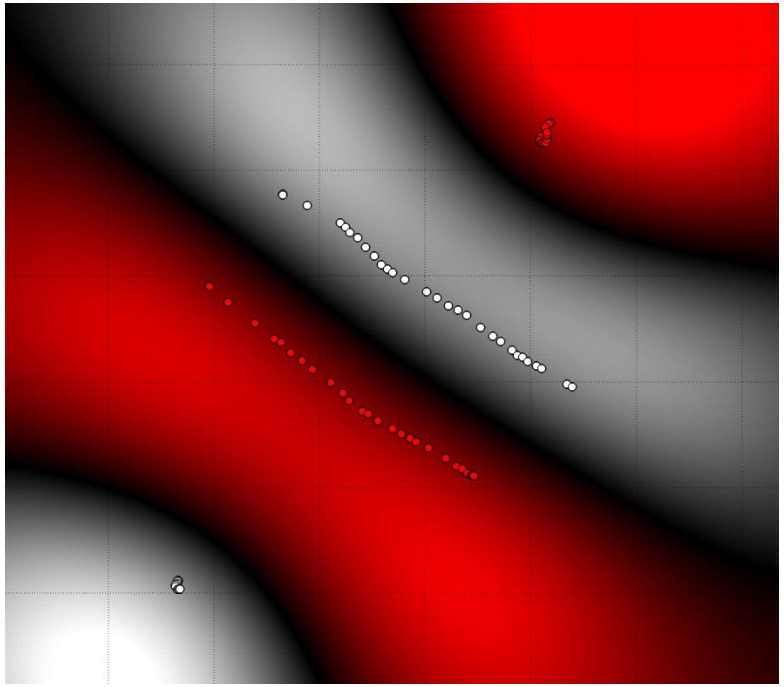
$\sigma = 0.01; C = 5000$



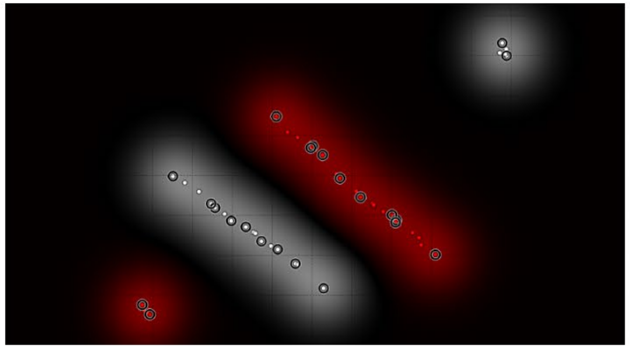
$\sigma = 0.5; C = 10$



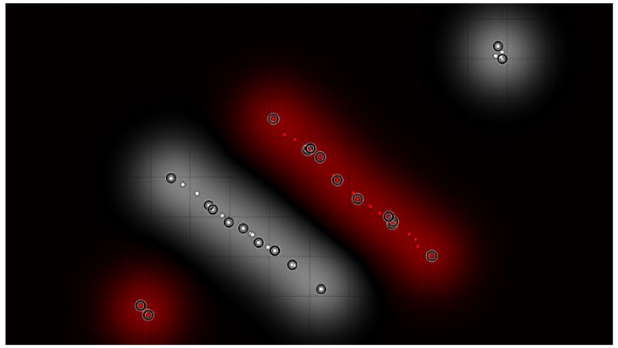
$\sigma = 0.1; C = 1000$



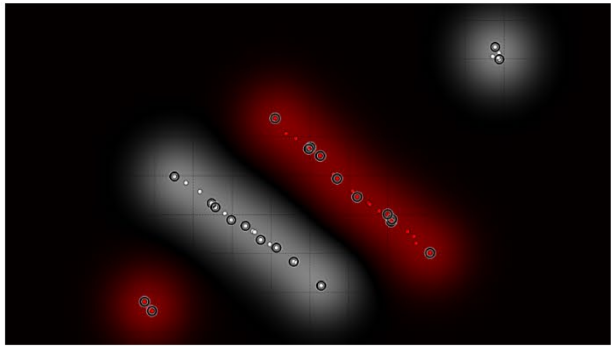
Q2 > C – Case ii



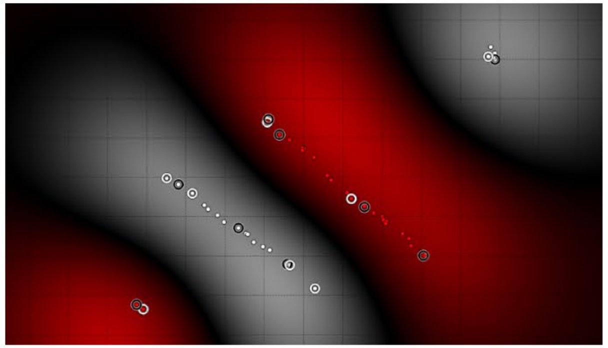
(a) $\sigma = 0.01; C = 1$



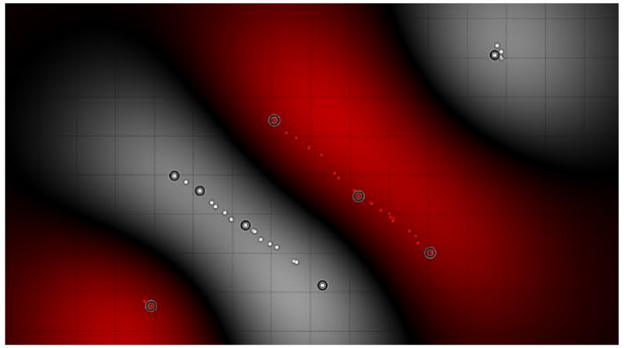
(b) $\sigma = 0.01; C = 10$



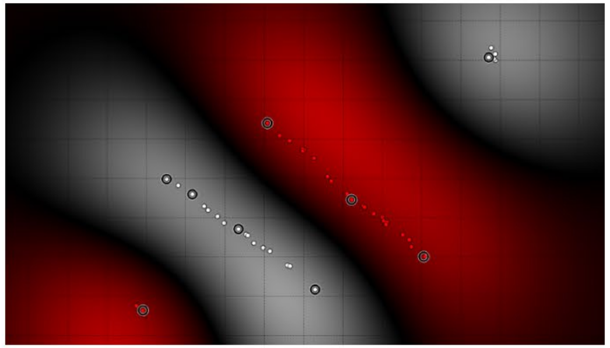
(c) $\sigma = 0.01; C = 5000$



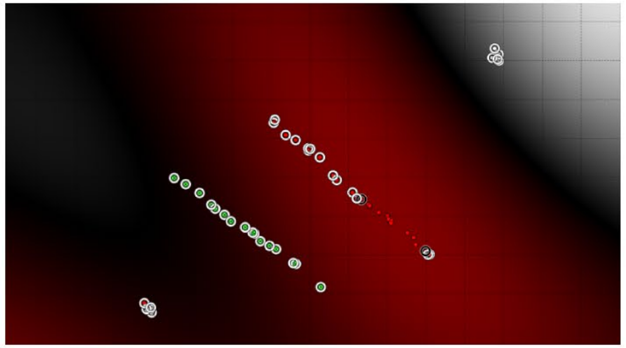
(d) $\sigma = 0.1; C = 1$



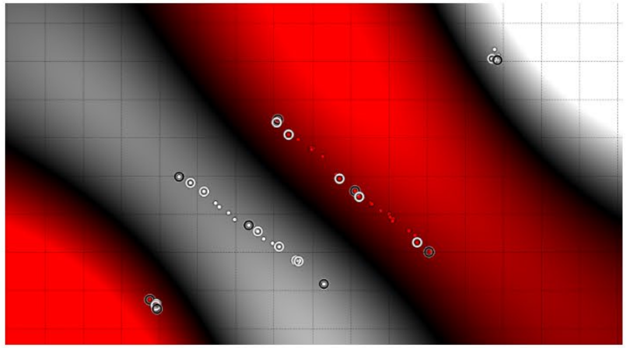
(e) $\sigma = 0.1; C = 10$



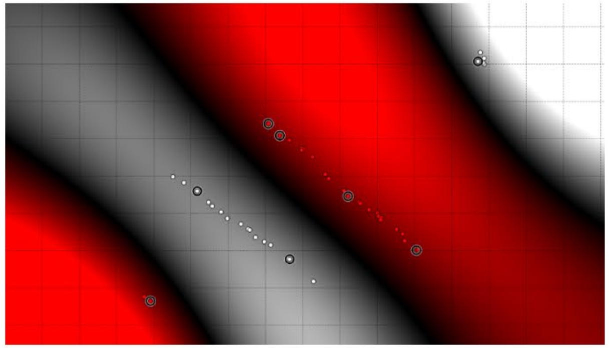
(f) $\sigma = 0.1; C = 5000$



(g) $\sigma = 0.5; C = 1$



(h) $\sigma = 0.5; C = 10$



(i) $\sigma = 0.5; C = 5000$