



EFC3 - Exercise 3

Rafael Claro Ito (R.A.: 118430)

November 2019

1 Source files

All code cited and all figures showed here can be found at the following GitHub repository:
<https://github.com/ito-rafael/IA006C-MachineLearning/tree/master/efc3>

In this repository, one can found the following files:

- Jupyter Notebook
 - MLP.ipynb
 - SVM.ipynb
- L^AT_EX
 - efc3.tex

The notebook “MLP” implements a multi-layer perceptron network used for binary classification. Here, we use different numbers of neurons, plot the decision regions and also calculate some metrics to evaluate the overall performance.

The notebook “SVM” uses the same dataset and also implements a binary classifier, but it does this using a support vector machine. Here, we plot the decision regions as well, but this time we vary the value of the hyperparameter C and also the value gamma used in a RBF kernel.

2 Part I - Error backpropagation

$$\begin{aligned}u_1 &= 1 \cdot v_{00} + x_1 \cdot v_{10} + x_2 \cdot v_{20} \\u_2 &= 1 \cdot v_{01} + x_1 \cdot v_{11} + x_2 \cdot v_{21} \\u_3 &= 1 \cdot v_{02} + x_1 \cdot v_{12} + x_2 \cdot v_{22}\end{aligned}$$

$$\begin{aligned}s_1 &= f(u_1) \\s_2 &= f(u_2) \\s_3 &= f(u_3)\end{aligned}$$

$$\begin{aligned}y_1 &= 1 \cdot w_{00} + s_1 \cdot w_{10} + s_2 \cdot w_{20} + s_3 \cdot w_{30} \\y_2 &= 1 \cdot w_{01} + s_1 \cdot w_{11} + s_2 \cdot w_{21} + s_3 \cdot w_{31}\end{aligned}$$

$$J = e_1^2 + e_2^2$$

$$\begin{aligned}\delta_3 &= \frac{\partial J}{\partial u_3} = \frac{\partial[(d_1 - y_1)^2 + (d_2 - y_2)^2]}{\partial u_3} \\ \delta_3 &= \frac{\partial(d_1 - y_1)^2}{\partial u_3} + \frac{\partial(d_2 - y_2)^2}{\partial u_3} \\ \delta_3 &= \frac{\partial(d_1 - y_1)^2}{\partial(d_1 - y_1)} \cdot \frac{\partial(d_1 - y_1)}{\partial u_3} + \frac{\partial(d_2 - y_2)^2}{\partial(d_2 - y_2)} \cdot \frac{\partial(d_2 - y_2)}{\partial u_3} \\ \delta_3 &= 2(d_1 - y_1) \cdot \left(-\frac{\partial y_1}{\partial u_3}\right) + 2(d_2 - y_2) \cdot \left(-\frac{\partial y_2}{\partial u_3}\right) \\ \delta_3 &= -2(d_1 - y_1) \cdot \frac{\partial y_1}{\partial s_3} \cdot \frac{\partial s_3}{\partial u_3} - 2(d_2 - y_2) \cdot \frac{\partial y_2}{\partial s_3} \cdot \frac{\partial s_3}{\partial u_3} \\ \delta_3 &= -2(d_1 - y_1) \cdot \frac{\partial(1 \cdot w_{00} + s_1 \cdot w_{10} + s_2 \cdot w_{20} + s_3 \cdot w_{30})}{\partial s_3} \cdot \frac{\partial f(u_3)}{\partial u_3} \\ &\quad - 2(d_2 - y_2) \cdot \frac{\partial(1 \cdot w_{01} + s_1 \cdot w_{11} + s_2 \cdot w_{21} + s_3 \cdot w_{31})}{\partial s_3} \cdot \frac{\partial f(u_3)}{\partial u_3}\end{aligned}$$

$$\delta_3 = -2(d_1 - y_1) \dot{f}(u_3) w_{30} - 2(d_2 - y_2) \dot{f}(u_3) w_{31}$$

$$\begin{aligned}\frac{\partial J}{\partial v_{12}} &= \frac{\partial J}{\partial u_3} \cdot \frac{\partial u_3}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{12}} &= \frac{\delta_3 \cdot \partial(1 \cdot v_{02} + x_1 \cdot v_{12} + x_2 \cdot v_{22})}{\partial v_{12}} \\ \frac{\partial J}{\partial v_{12}} &= \delta_3 \cdot x_1 \\ \frac{\partial J}{\partial v_{12}} &= -2x_1 f(u_3) [w_{30}(d_1 - y_1) + w_{31}(d_2 - y_2)]\end{aligned}$$

3 Part II - Binary Classification with MLP and SVMs

3.1 Multi-layer Perceptron (MLP)

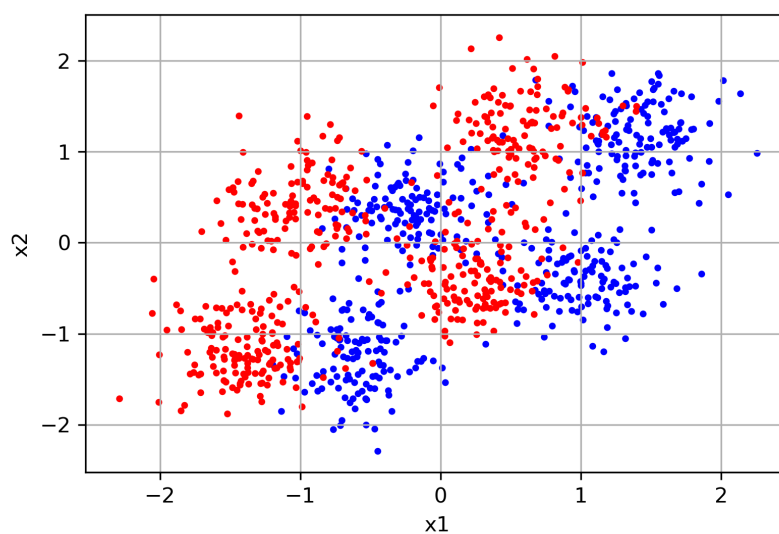


Figure 2

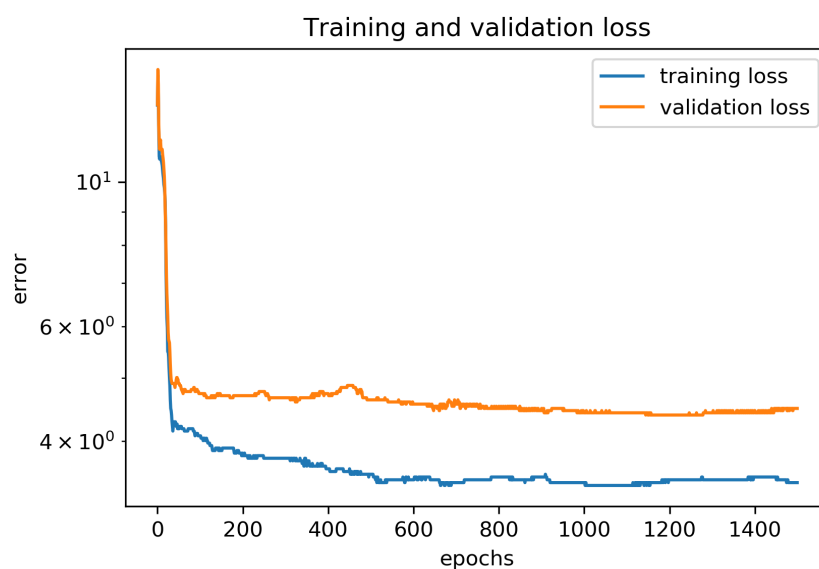


Figure 3

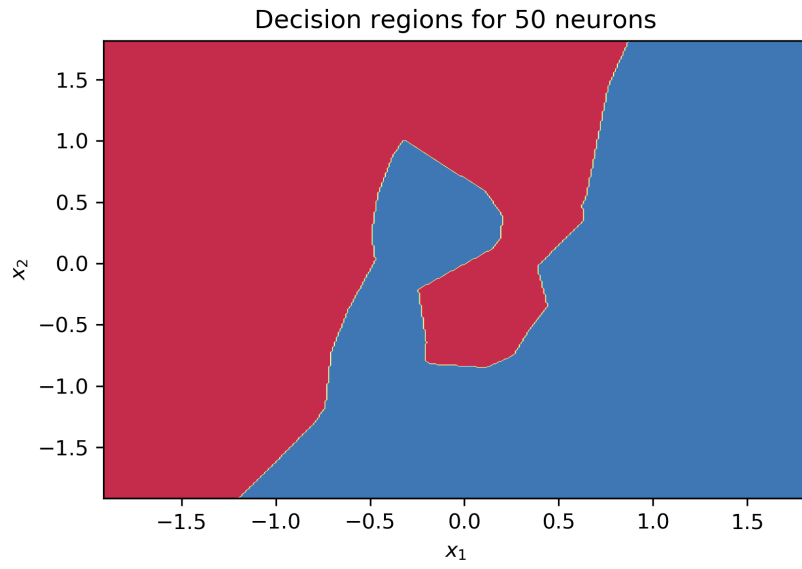


Figure 4

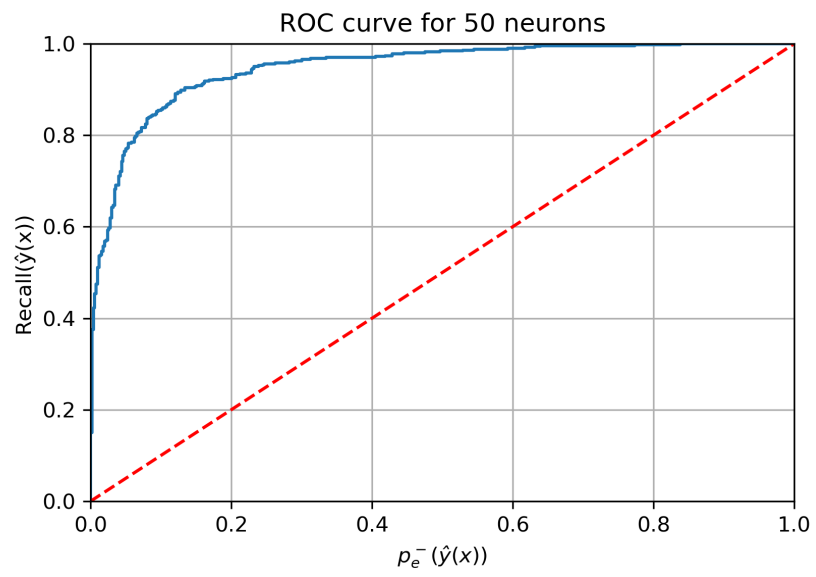


Figure 5

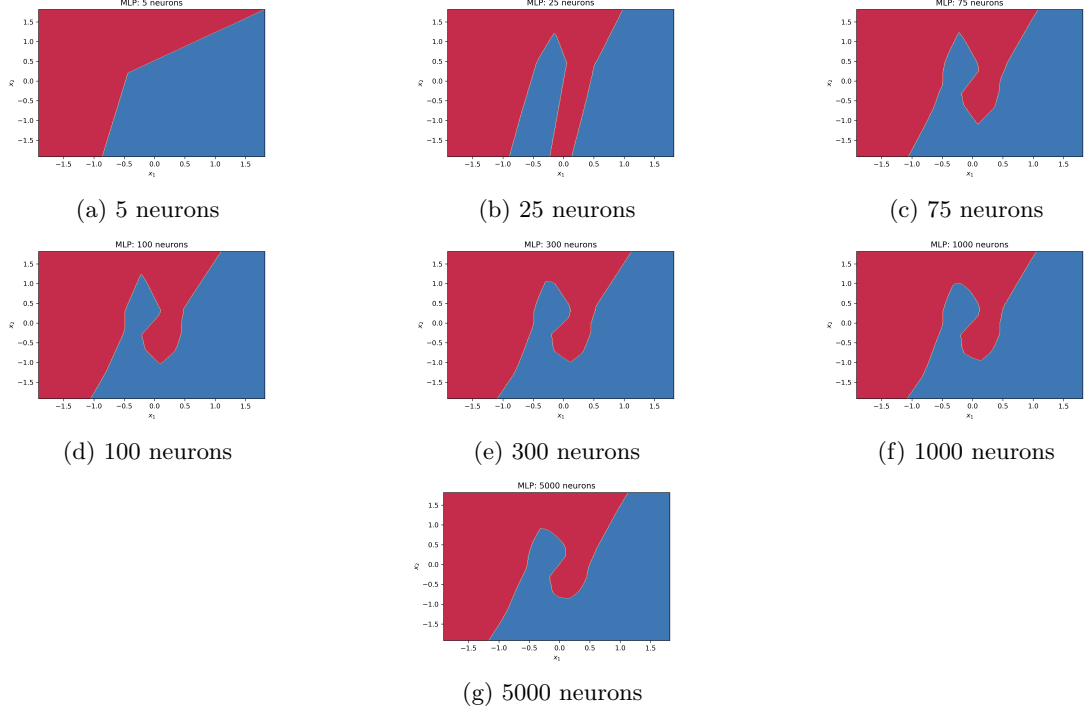


Figure 6: Decision regions for MLP with different number of neurons

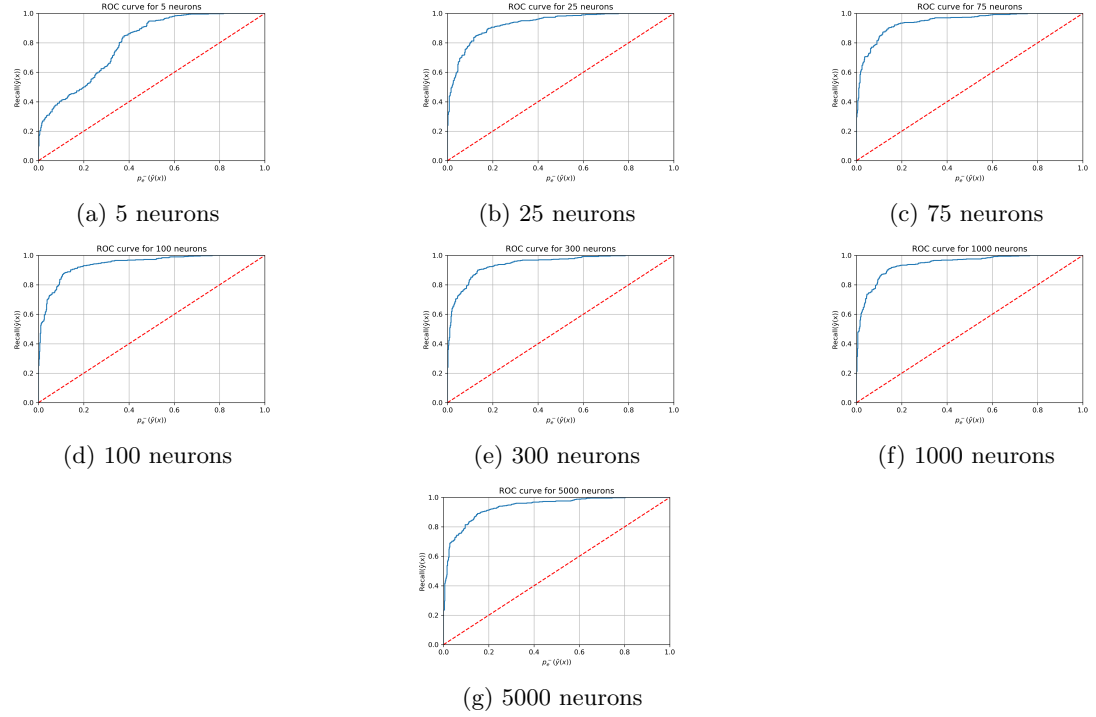


Figure 7: ROC curves for MLP with different number of neurons

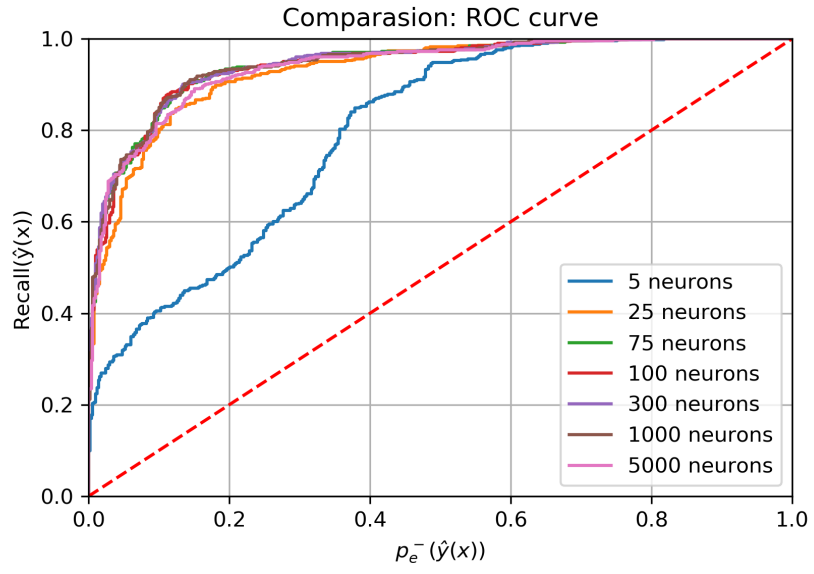


Figure 8

3.2 Support Vector Machine (SVM)

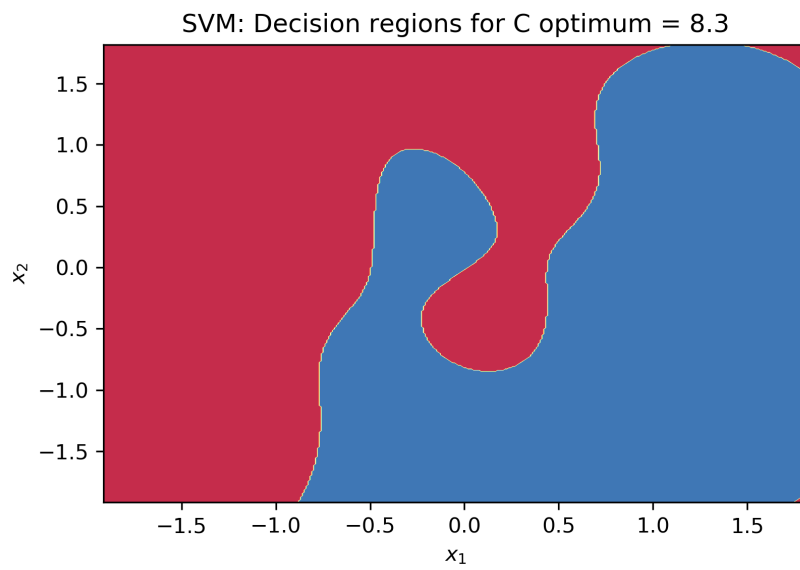


Figure 9: Decision regions

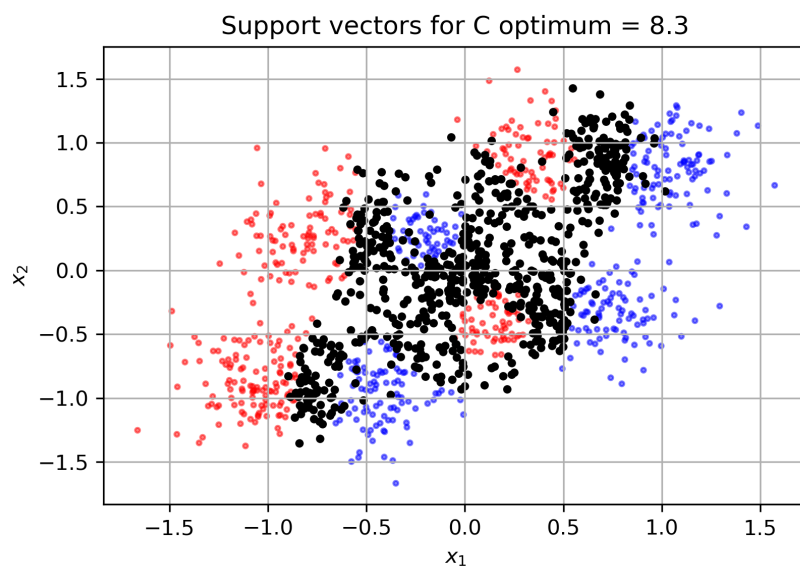
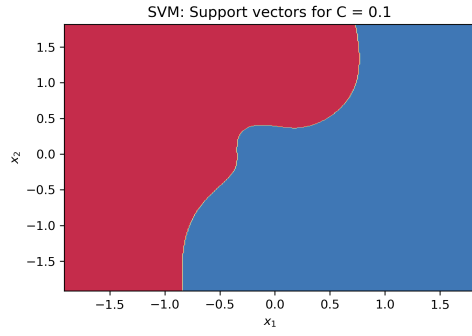
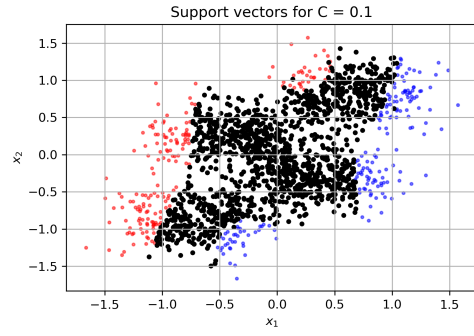


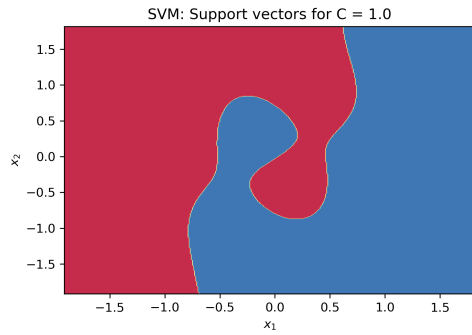
Figure 10: Support vectors



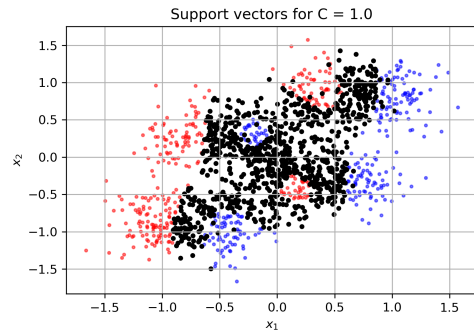
(a) Decision regions for $C = 0.1$



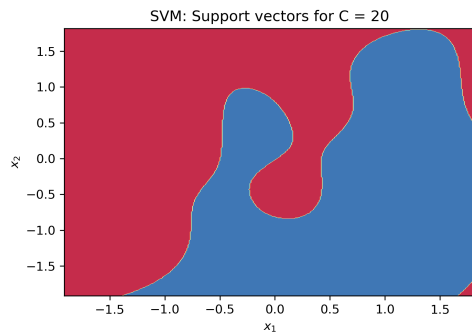
(b) Support vectors for $C = 0.1$



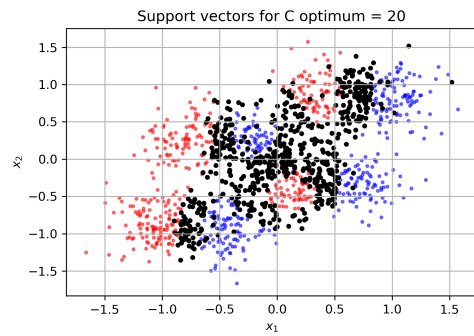
(c) Decision regions for $C = 1.0$



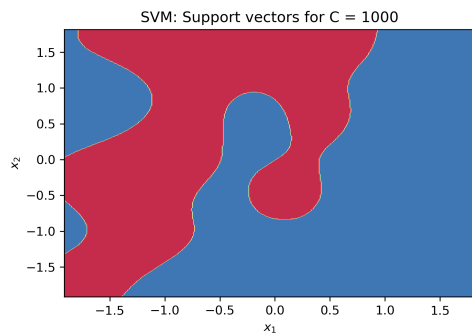
(d) Support vectors for $C = 1.0$



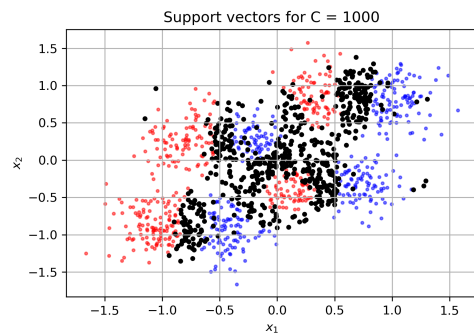
(e) Decision regions for $C = 20$



(f) Support vectors for $C = 20$

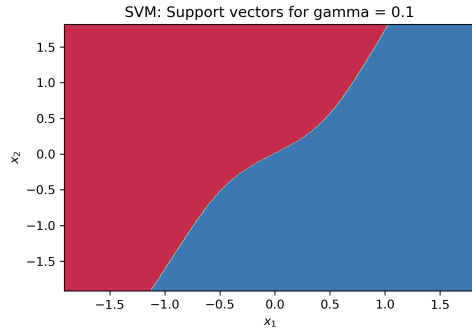


(g) Decision regions for $C = 1000$

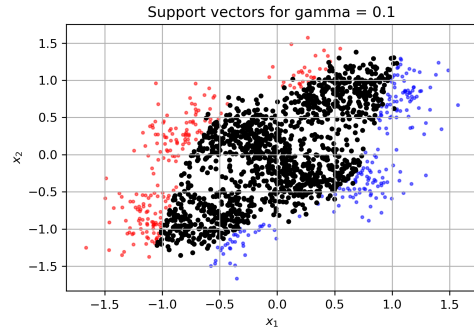


(h) Support vectors for $C = 1000$

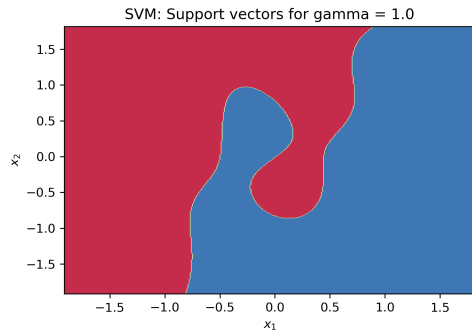
Figure 11: Decision regions and respective support vectors for different values of hyperparameter C



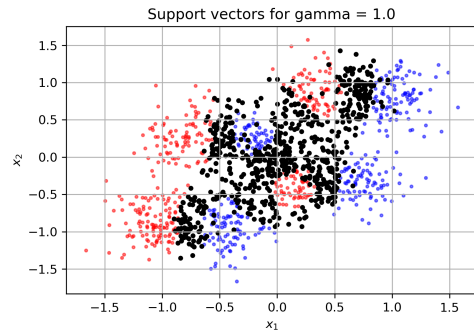
(a) Decision regions for gamma = 0.1



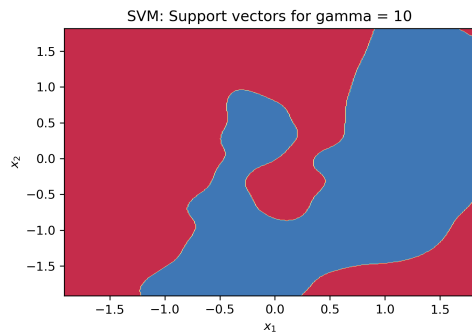
(b) Support vectors for gamma = 0.1



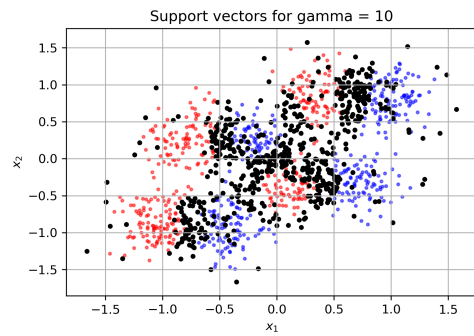
(c) Decision regions for gamma = 1.0



(d) Support vectors for gamma = 1.0



(e) Decision regions for gamma = 10



(f) Support vectors for gamma = 10

Figure 12: Decision regions and respective support vectors for different values of hyperparameter gamma in kernel function