

SPRAWOZDANIE 5

Podstawy Sztucznej Inteligencji

Klasyfikator SVM

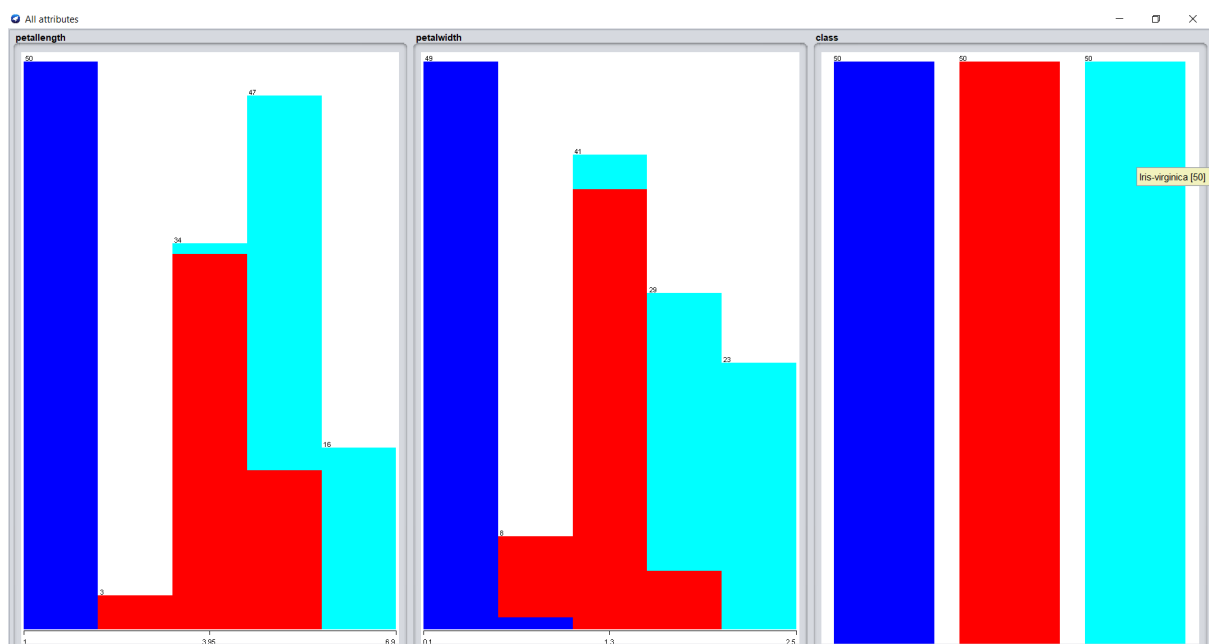
Program WEKA 5

Natalia Gadocha 304165
Geoinformatyka III rok

Naszym obiektem badań są dane iris2D. Będziemy korzystać z klasyfikatora SVM. SVM'y służą do klasyfikacji binarnej.

O metodzie SVM można ogólnie powiedzieć, że:

- w przypadku liniowo separowalnym, tj. wtedy, gdy istnieje przynajmniej jedna płaszczyzna separacji oddzielająca klasy, metoda ta gwarantuje znalezienie takiej płaszczyzny, która ma maksymalny tzw. margines separacji;
- w przypadku nieseparowalnym liniowo, metoda SVM pozwala na znalezienie płaszczyzny, która klasyfikuje obiekty na tyle poprawnie, na ile jest to możliwe i jednocześnie przebiega możliwie daleko od typowych skupień dla każdej z klas (będziemy tu mówili o największym marginesie w sensie pewnej zadanej heurystyki);
- w przypadku nieseparowalnym liniowo, stosując tzw. podniesienie wymiarowości, można za pomocą metody SVM znaleźć krzywoliniową granicę klasyfikacji o dużym marginesie separacji.



LibLINEAR - classify

Correctly Classified Instances	139	92.6667 %
Incorrectly Classified Instances	11	7.3333 %
Kappa statistic	0.89	
Mean absolute error	0.0489	
Root mean squared error	0.2211	
Relative absolute error	11 %	
Root relative squared error	46.9042 %	
Total Number of Instances	150	
Precision	0,926	

```
a b c <-- classified as
50 0 0 | a = Iris-setosa
1 44 5 | b = Iris-versicolor
0 5 45 | c = Iris-virginica
```

LibSVM - classify

Correctly Classified Instances	145	96.6667 %
Incorrectly Classified Instances	5	3.3333 %
Kappa statistic	0.95	
Mean absolute error	0.0222	
Root mean squared error	0.1491	
Relative absolute error	5 %	
Root relative squared error	31.6228 %	
Total Number of Instances	150	
Precision	0,967	

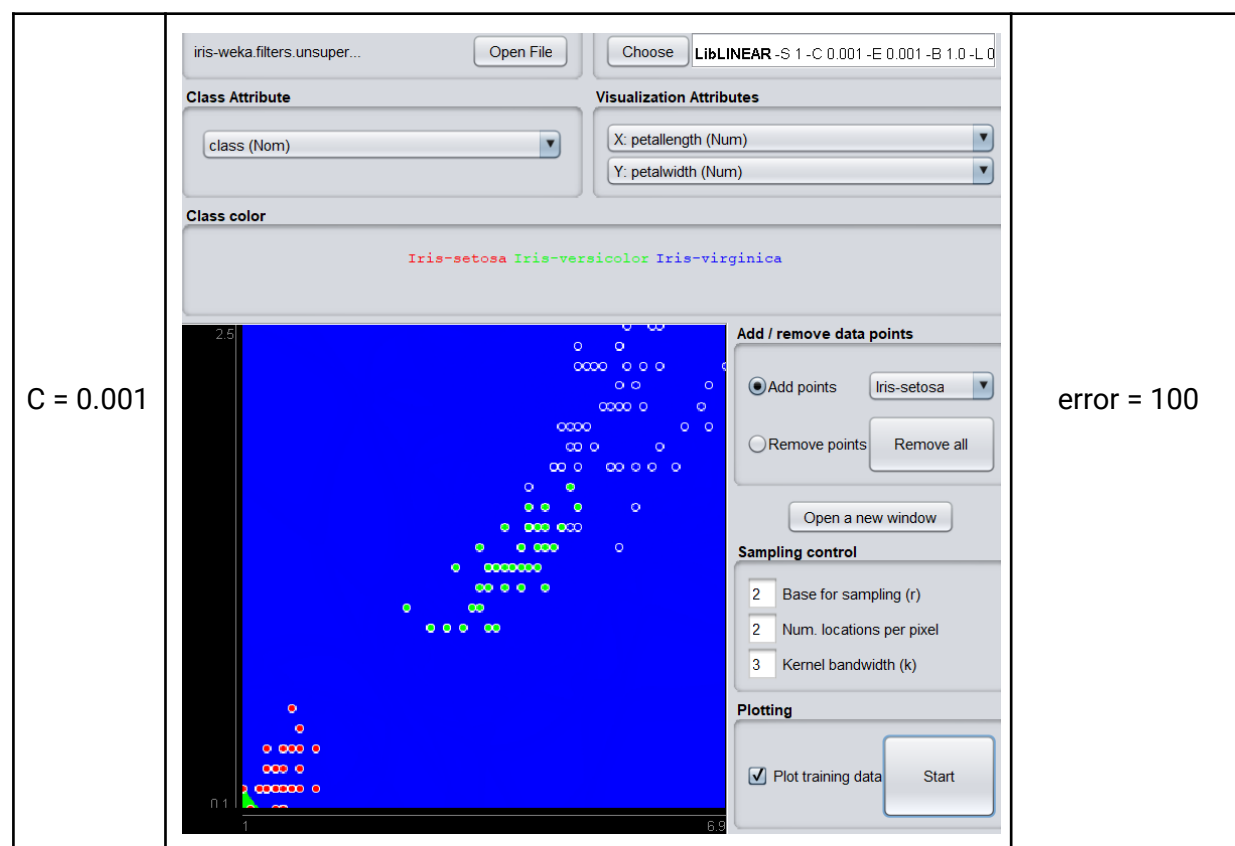
```
a b c <-- classified as
50 0 0 | a = Iris-setosa
0 47 3 | b = Iris-versicolor
0 2 48 | c = Iris-virginica
```

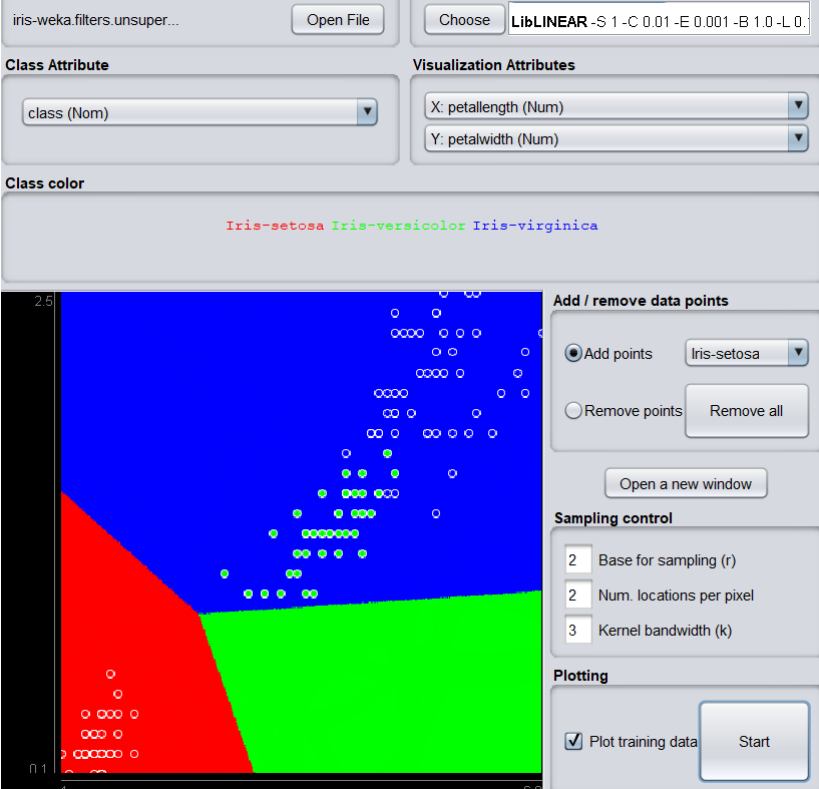
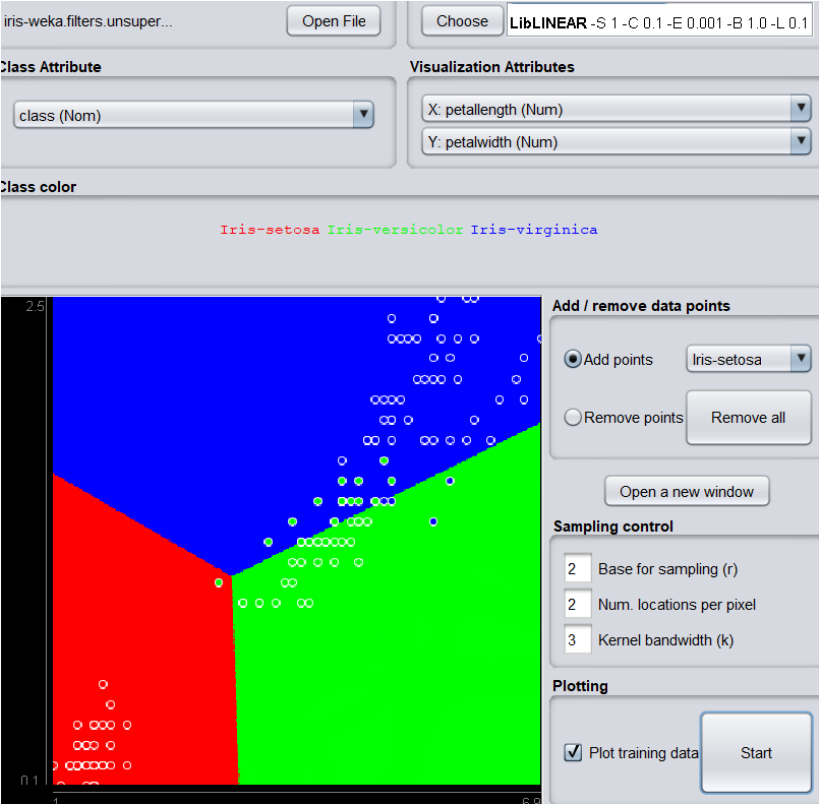
Na podstawie samych uzyskanych powyższych danych możemy dostrzec większe zalety klasyfikatora LibSVM. Ma zdecydowanie większą poprawną trafność przyporządkowania, niewielkie błędy oraz wysokie pozostałe czynniki. Jego precyzja wynosi 0,967, co jest bardzo wysoką wartością. LibLINEAR również wypada w tym zestawieniu naprawdę wysoko, jednakże jego wyniki są słabsze.

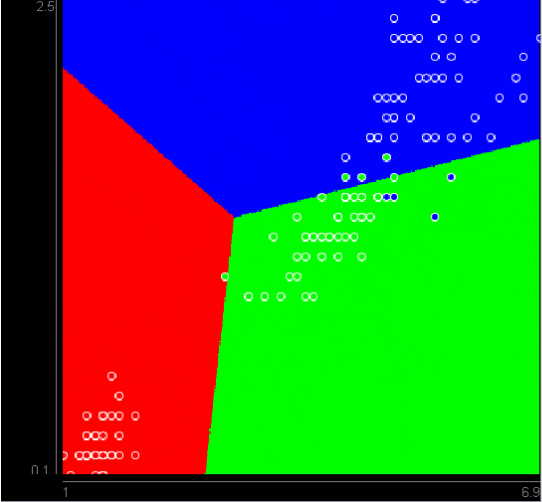
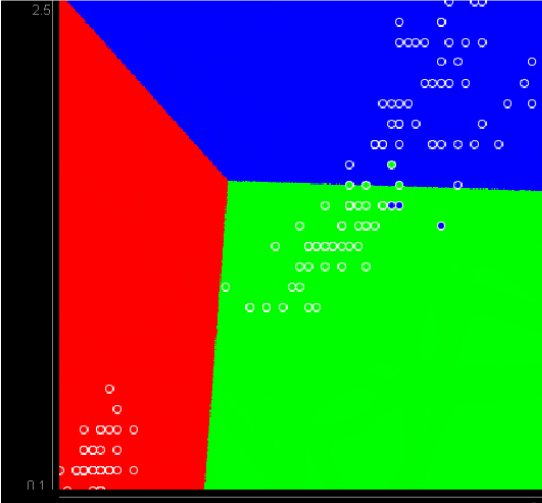
Wizualizacja danych

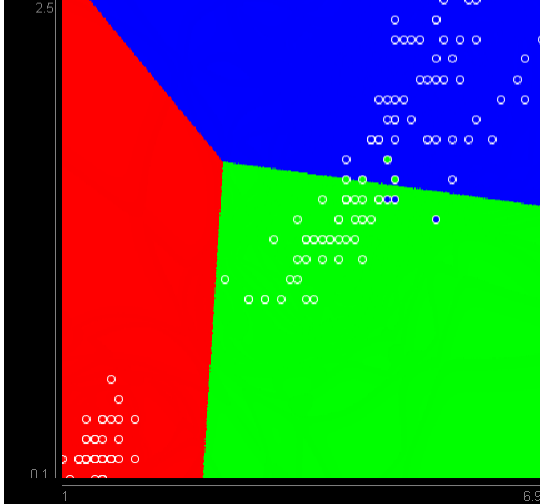
Zakres danych z jakiego będziemy korzystać to: C -> 0.001 do 100 oraz gamma -> 0.001 do 100

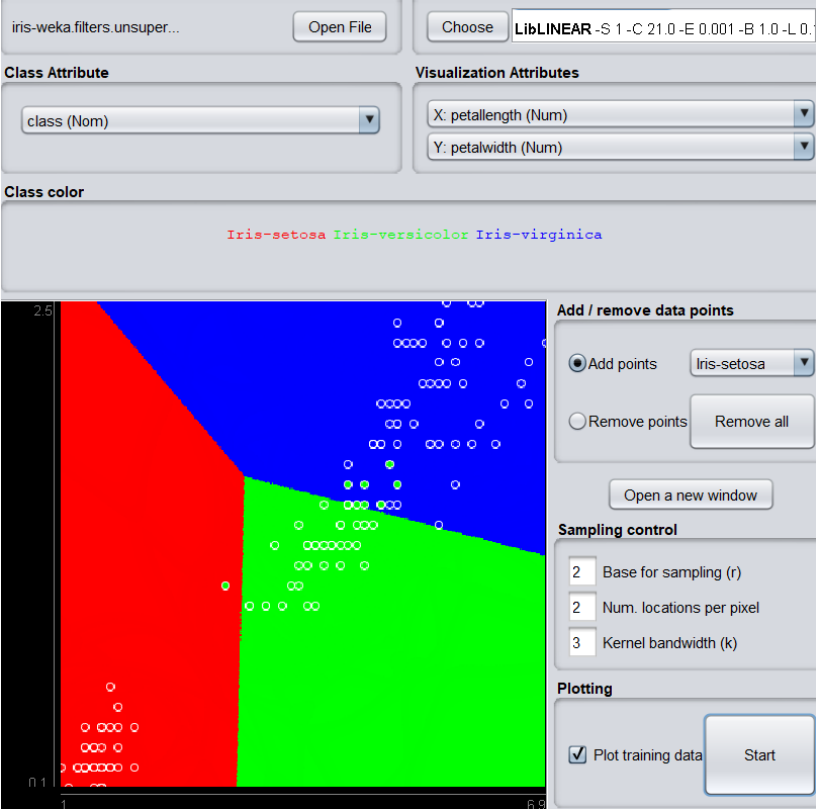
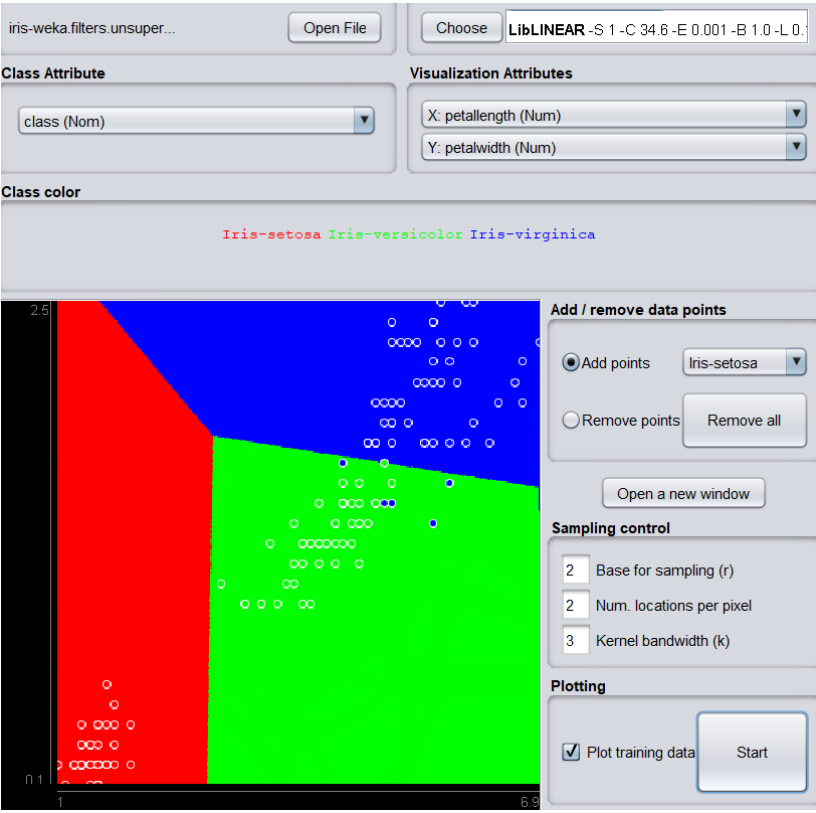
LibLINEAR

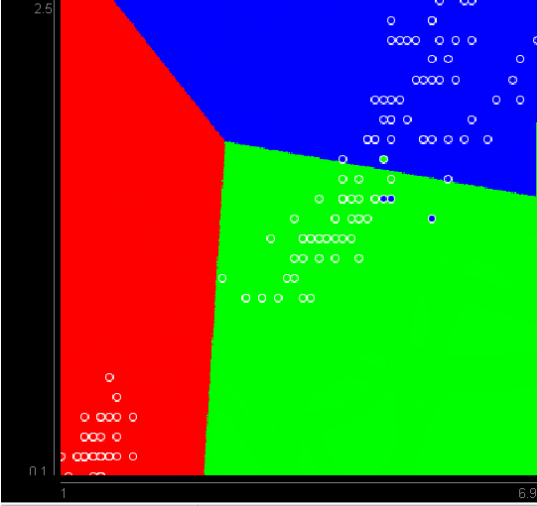


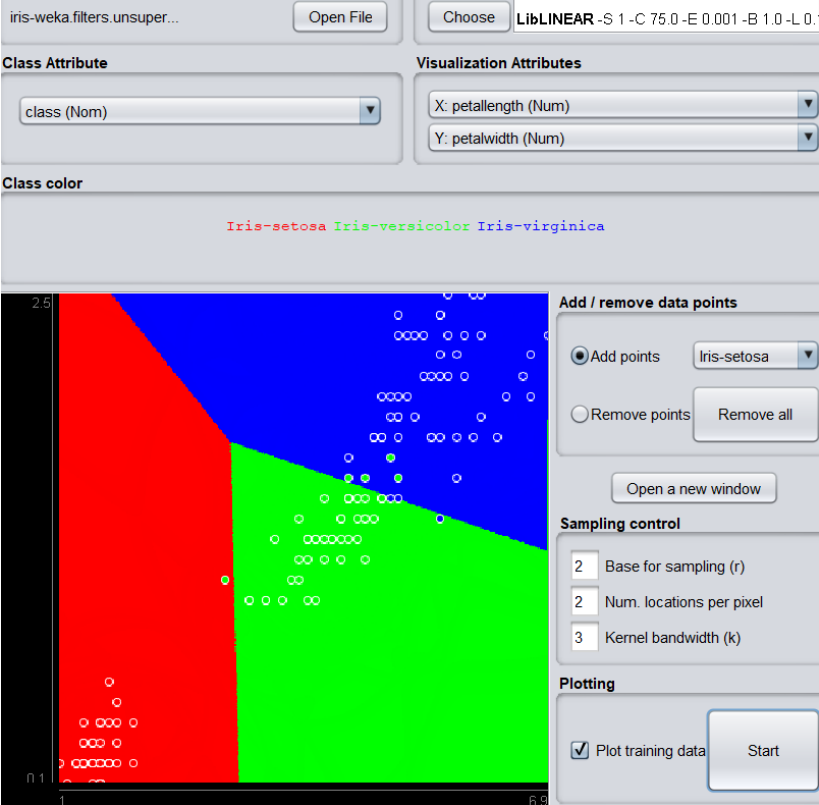
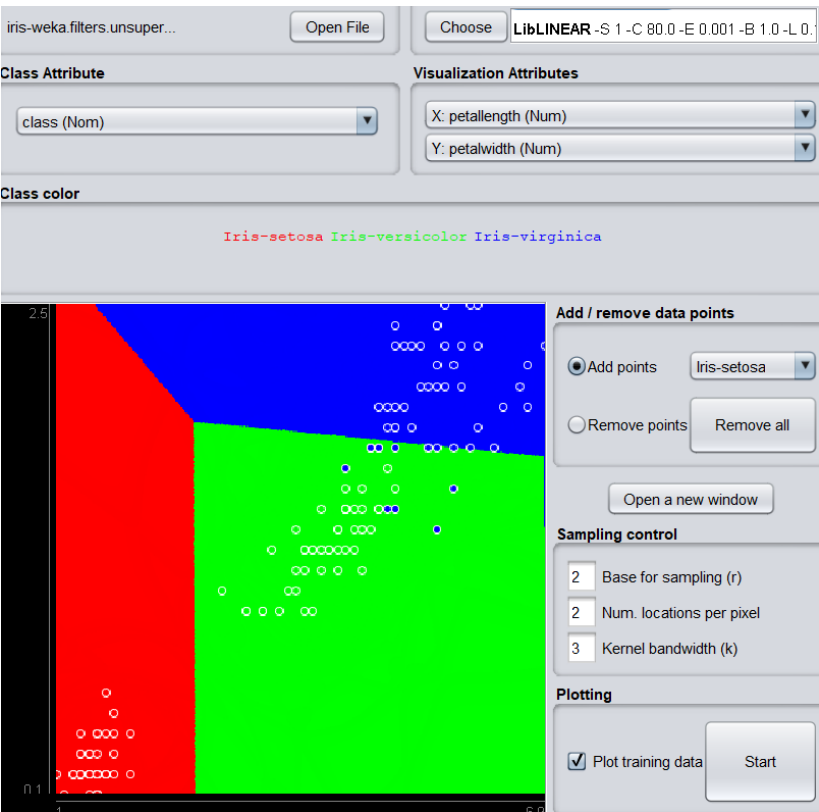
<p>C = 0.01</p>		<p>error = 50</p>
<p>C = 0.1</p>		<p>error = 18</p>

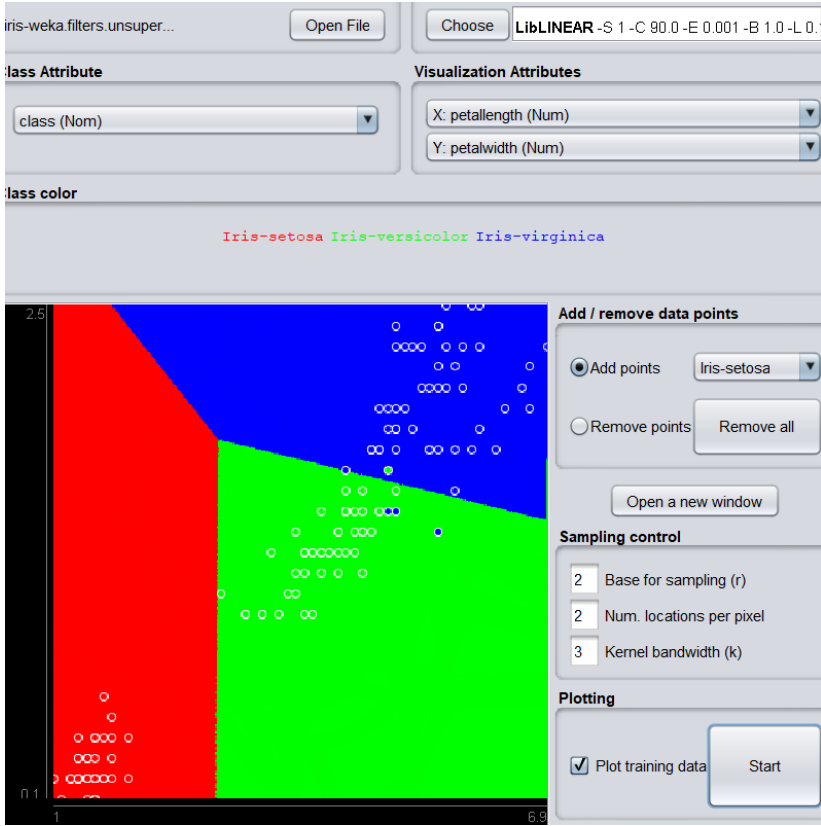
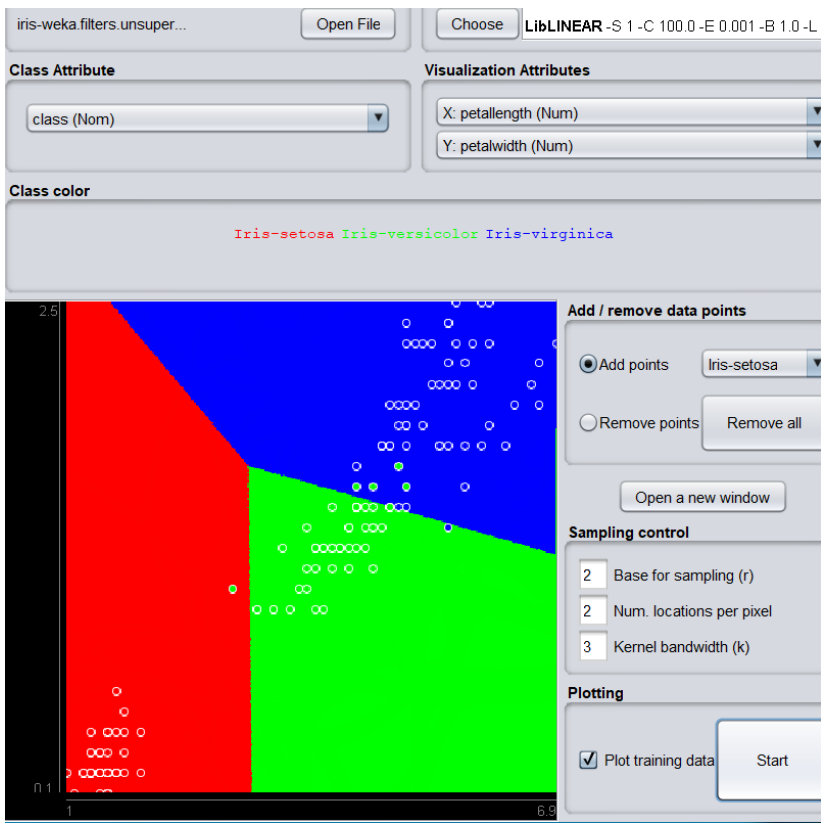
<p>C = 1</p>	<div data-bbox="359 210 1182 1016"> <div>iris-weka.filters.unsuper... Open File</div> <div>Choose LibLINEAR -S 1 -C 1.0 -E 0.001 -B 1.0 -L 0.1</div> <div> <div>Class Attribute</div> <div>class (Nom)</div> </div> <div> <div>Visualization Attributes</div> <div>X: petallength (Num)</div> <div>Y: petalwidth (Num)</div> </div> <div> <div>Class color</div> <div>Iris-setosa Iris-versicolor Iris-virginica</div> </div> <div>  <div> <div>Add / remove data points</div> <div> <input checked="" type="radio"/> Add points <div>Iris-setosa</div> </div> <div> <input type="radio"/> Remove points <div>Remove all</div> </div> <div>Open a new window</div> </div> <div> <div>Sampling control</div> <div>2 Base for sampling (r)</div> <div>2 Num. locations per pixel</div> <div>3 Kernel bandwidth (k)</div> </div> <div> <div>Plotting</div> <div><input checked="" type="checkbox"/> Plot training data</div> <div>Start</div> </div> </div> </div>	<p>error = 10</p>
<p>C = 5</p>	<div data-bbox="359 1048 1182 1854"> <div>iris-weka.filters.unsuper... Open File</div> <div>Choose LibLINEAR -S 1 -C 5.0 -E 0.001 -B 1.0 -L 0.1</div> <div> <div>Class Attribute</div> <div>class (Nom)</div> </div> <div> <div>Visualization Attributes</div> <div>X: petallength (Num)</div> <div>Y: petalwidth (Num)</div> </div> <div> <div>Class color</div> <div>Iris-setosa Iris-versicolor Iris-virginica</div> </div> <div>  <div> <div>Add / remove data points</div> <div> <input checked="" type="radio"/> Add points <div>Iris-setosa</div> </div> <div> <input type="radio"/> Remove points <div>Remove all</div> </div> <div>Open a new window</div> </div> <div> <div>Sampling control</div> <div>2 Base for sampling (r)</div> <div>2 Num. locations per pixel</div> <div>3 Kernel bandwidth (k)</div> </div> <div> <div>Plotting</div> <div><input checked="" type="checkbox"/> Plot training data</div> <div>Start</div> </div> </div> </div>	<p>error = 7</p>

<p>C = 10</p>	<div data-bbox="359 210 1182 1012"> <div>iris-weka.filters.unsuper... Open File</div> <div>Choose LibLINEAR -S 1 -C 10.0 -E 0.001 -B 1.0 -L 0.</div> <div> <div>Class Attribute</div> <div>class (Nom)</div> </div> <div> <div>Visualization Attributes</div> <div>X: petallength (Num)</div> <div>Y: petalwidth (Num)</div> </div> <div> <div>Class color</div> <div>Iris-setosa Iris-versicolor Iris-virginica</div> </div> <div>  </div> <div> <div>Add / remove data points</div> <div> <input checked="" type="radio"/> Add points Iris-setosa <input type="radio"/> Remove points Remove all </div> <div>Open a new window</div> </div> <div> <div>Sampling control</div> <div>2 Base for sampling (r)</div> <div>2 Num. locations per pixel</div> <div>3 Kernel bandwidth (k)</div> </div> <div> <div>Plotting</div> <div><input checked="" type="checkbox"/> Plot training data Start</div> </div> </div>
---------------	---

<p>C = 21</p>		<p>error = 9</p>
<p>C = 34.6</p>		<p>error = 6</p>

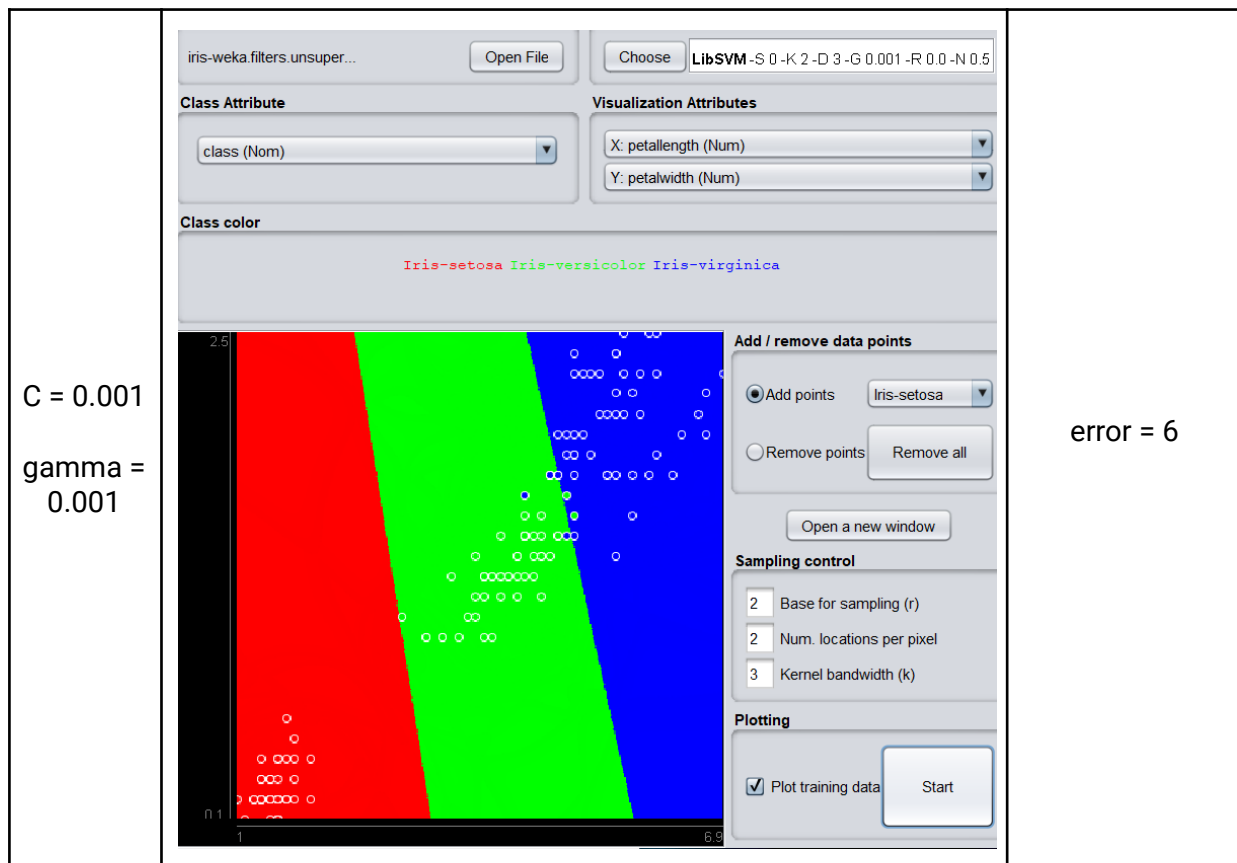
<p>C = 50</p>	<div data-bbox="363 210 1182 999"> <div>iris-weka.filters.unsuper... Open File</div> <div>Choose LibLINEAR -S 1 -C 50.0 -E 0.001 -B 1.0 -L 0.</div> <div> <div>Class Attribute</div> <div>class (Nom)</div> </div> <div> <div>Visualization Attributes</div> <div>X: petallength (Num)</div> <div>Y: petalwidth (Num)</div> </div> <div> <div>Class color</div> <div>Iris-setosa Iris-versicolor Iris-virginica</div> </div> <div>  </div> <div> <div>Add / remove data points</div> <div> <input checked="" type="radio"/> Add points <div>Iris-setosa</div> </div> <div> <input type="radio"/> Remove points <div>Remove all</div> </div> <div>Open a new window</div> </div> <div> <div>Sampling control</div> <div>2 Base for sampling (r)</div> <div>2 Num. locations per pixel</div> <div>3 Kernel bandwidth (k)</div> </div> <div> <div>Plotting</div> <div><input checked="" type="checkbox"/> Plot training data</div> <div>Start</div> </div> </div>
---------------	---

<p>C = 75</p>		<p>error = 8</p>
<p>C = 80</p>		<p>error = 13</p>

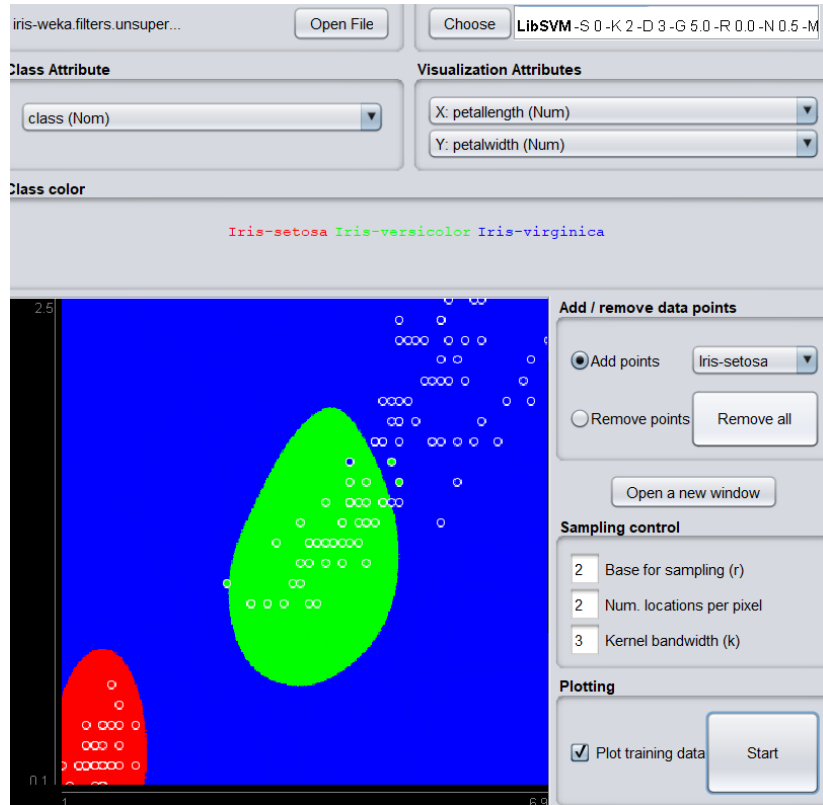
<p>C = 90</p>		<p>error = 6</p>
<p>C = 100</p>		<p>error = 7</p>

Na podstawie uzyskanych wyników możemy dostrzec, iż bardzo ciężko jest pozbyć się wszystkich wynikających błędów. Najmniejszy jaki udało mi się tu osiągnąć wynosił 4 dla parametru C równego 50. Nie zawsze też większa wartość C będzie się równała z mniejszymi błędami. Dla $C = 100$ ich ilość jest równa 50, a więc większa niż dla omawianego poprzednika. Największe błędy wystąpiły podczas najniższego C. Tam można było dostrzec aż 100 błędnych pomiarów. Największym problemem dla LibLINEAR było bliskie rozłożenie danych klas. Hiperpłaszczyzny jakie wyznacza tworzą linie proste, co dodatkowo utrudnia zadanie i poprawne przyporządkowanie.

LibSVM

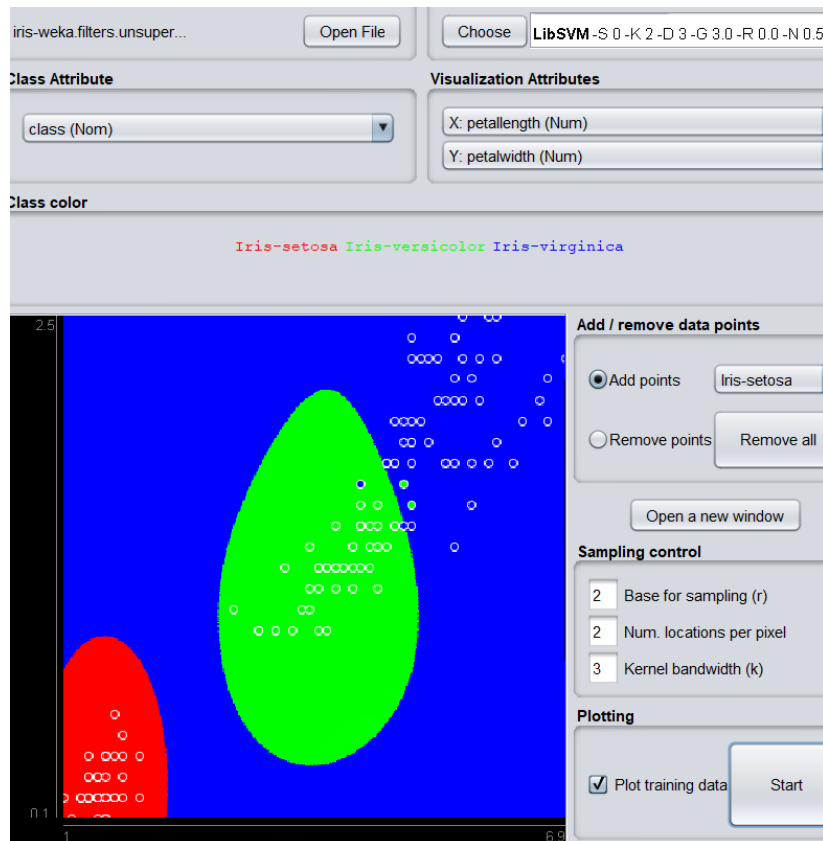


C = 0.001
gamma = 5



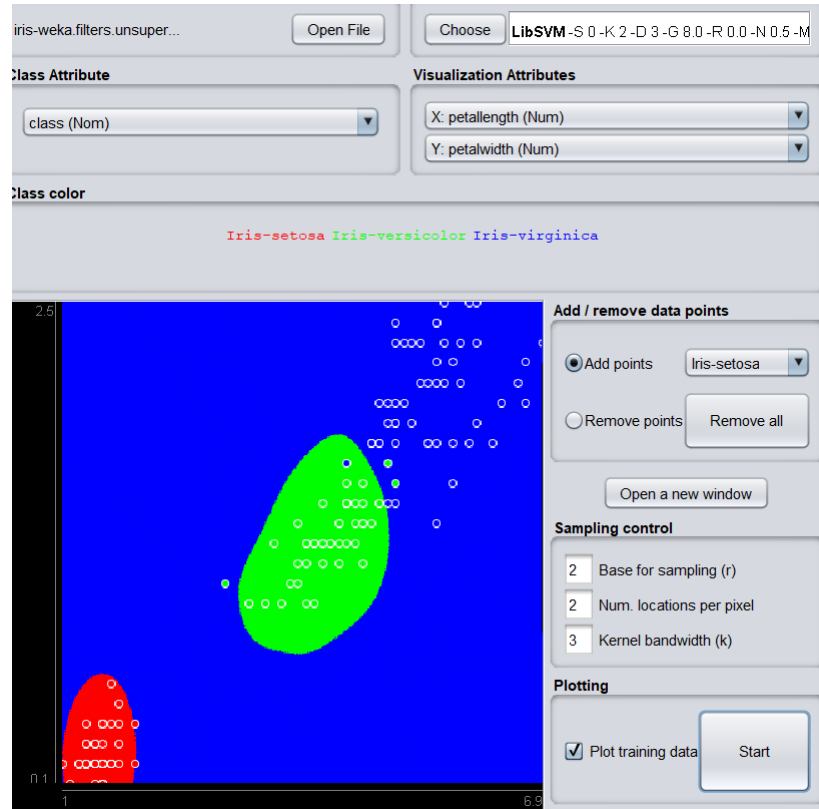
error = 4

C = 0.001
gamma = 3.0



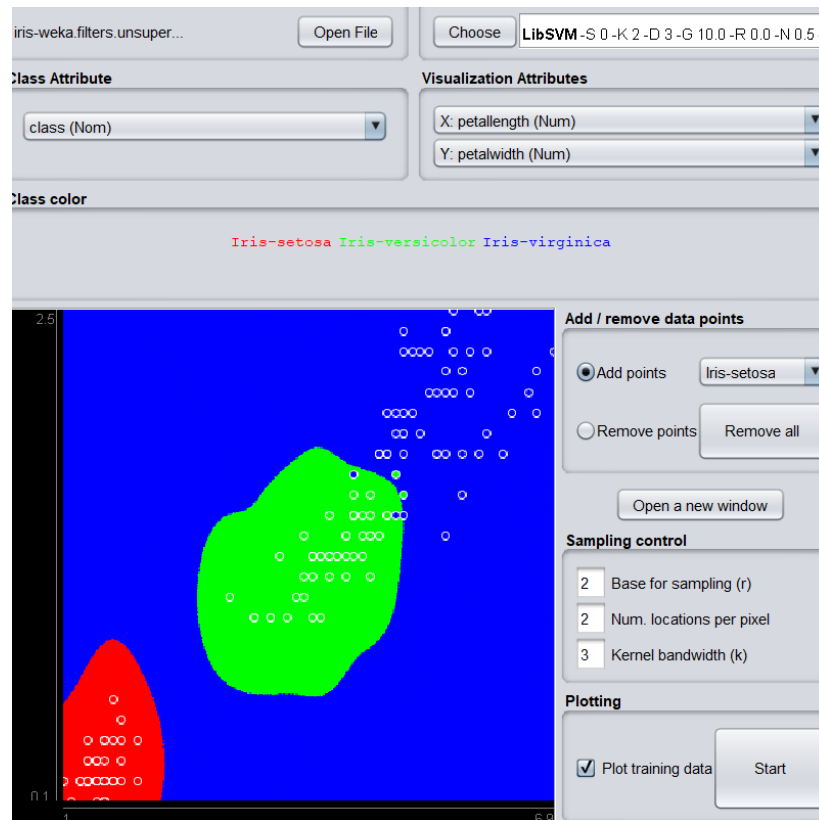
error = 4

C = 0.001
gamma = 8.0



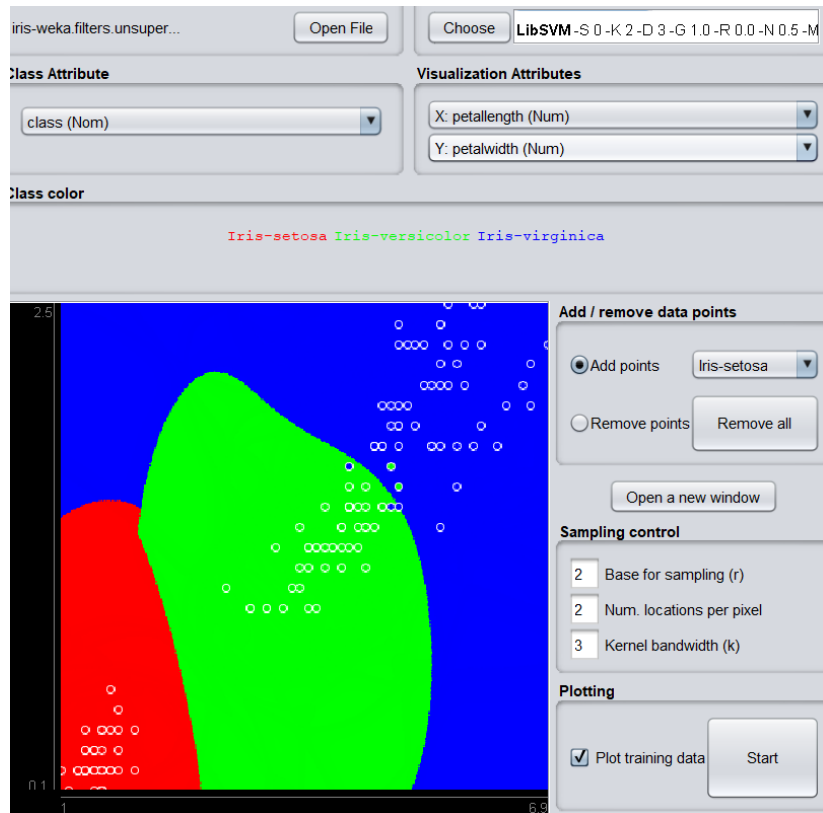
error = 6

C = 0,1
gamma = 10



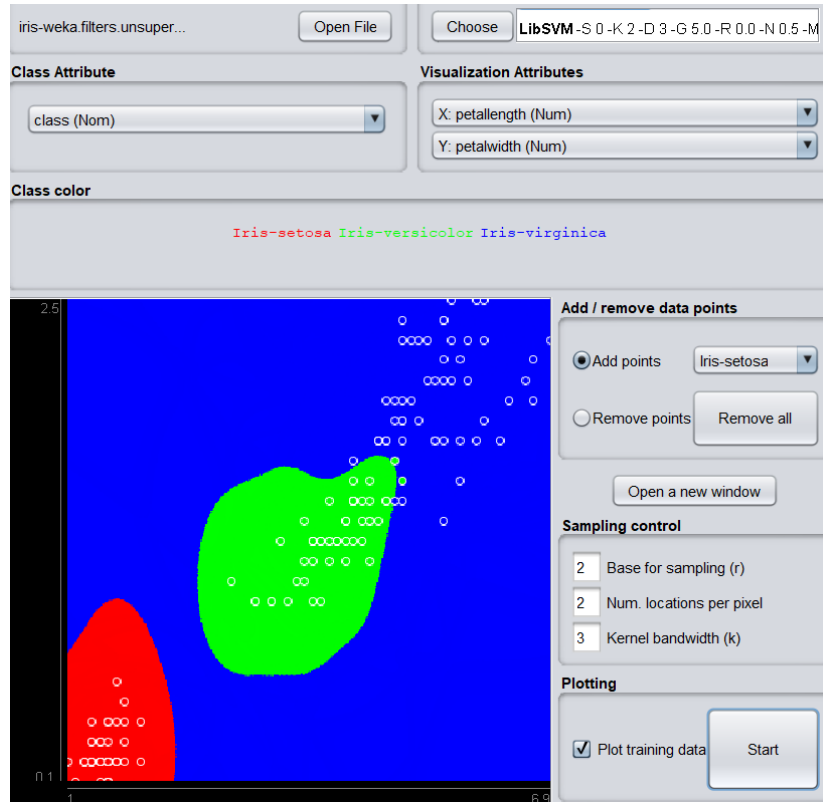
error = 5

C = 5
gamma = 1

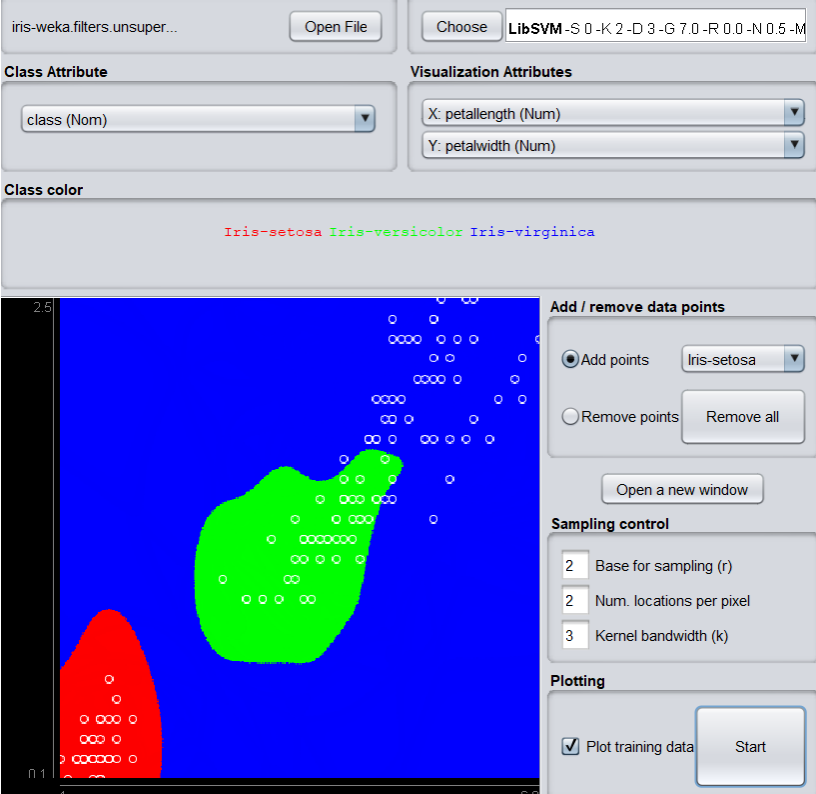
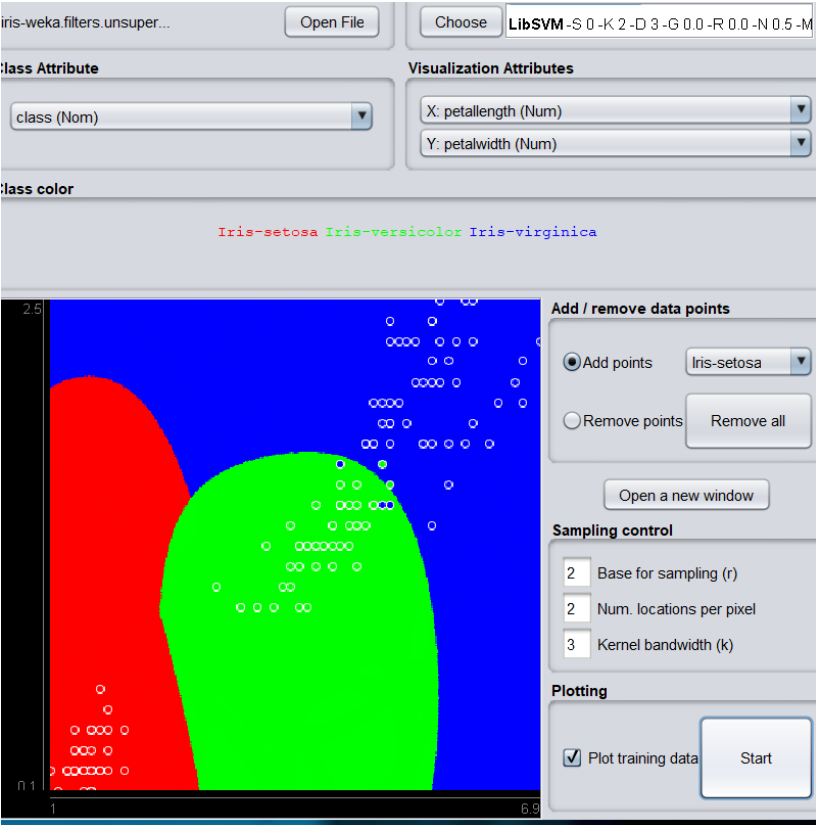


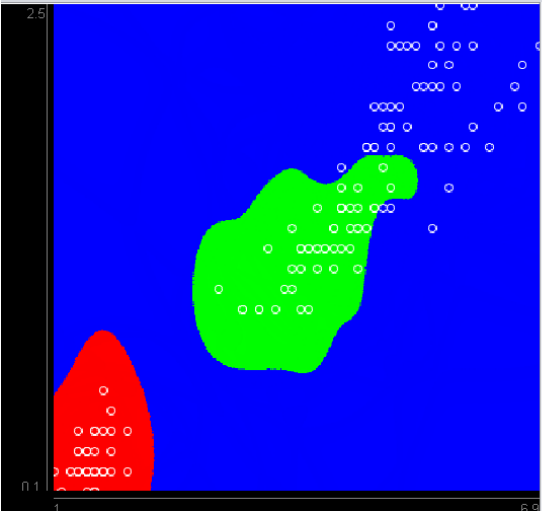
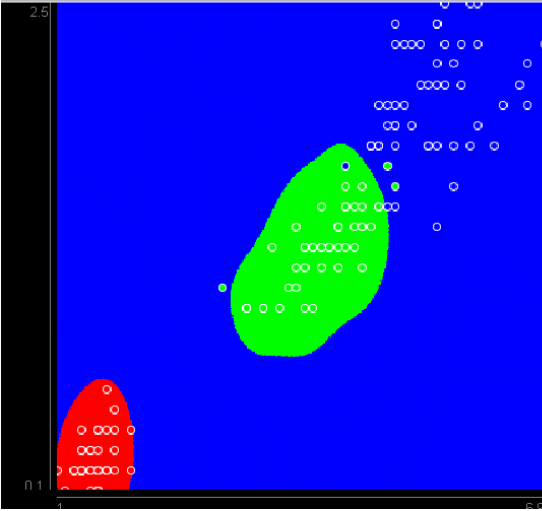
error = 5

C = 15
gamma = 2

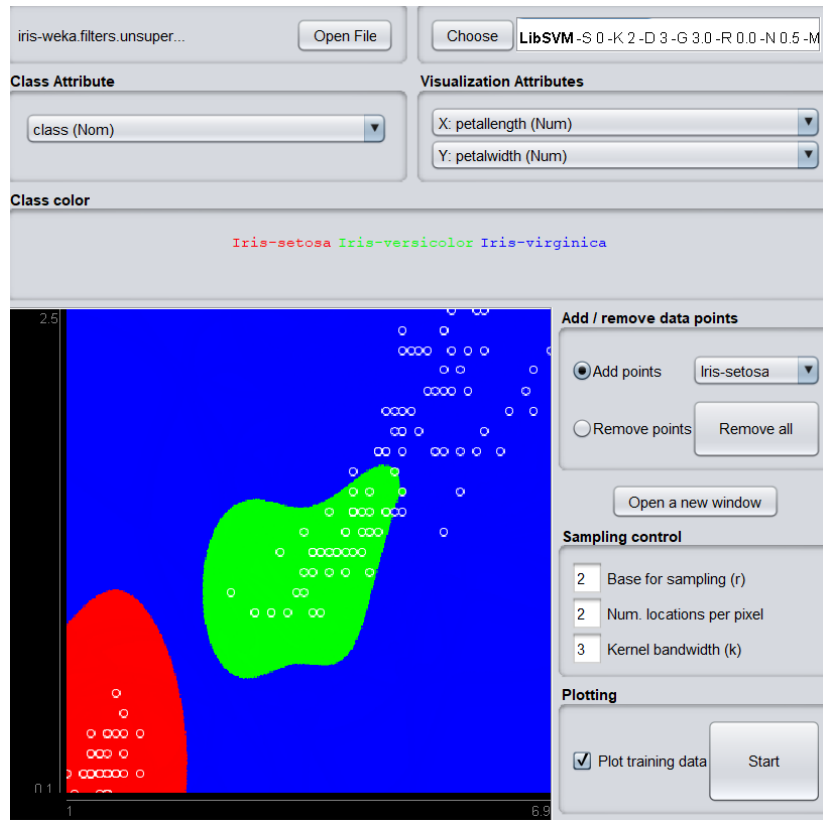


error = 3

<p> $C = 17$ $\gamma = 7$ </p>		<p>error = 0</p>
<p> $C = 20$ $\gamma = 0,009$ </p>		<p>error = 5</p>

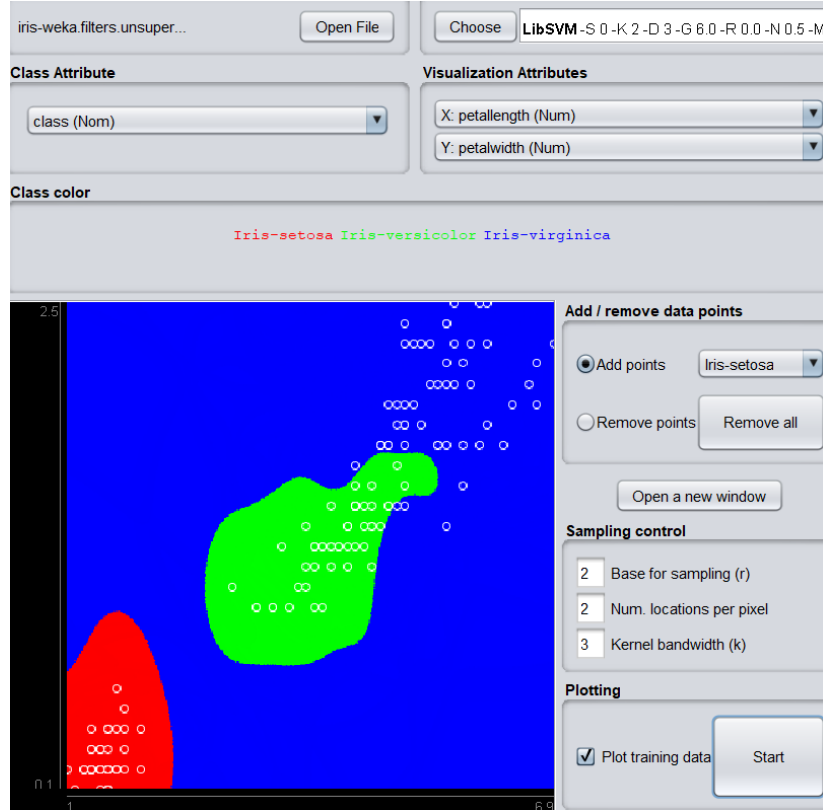
<p>C = 21</p> <p>gamma = 10</p>	<div><div>iris-weka.filters.unsuper... Open File</div><div>Choose LibSVM -S 0 -K 2 -D 3 -G 10.0 -R 0.0 -N 0.5 -</div><div>Class Attribute class (Nom)</div><div>Visualization Attributes X: petallength (Num) Y: petalwidth (Num)</div><div>Class color Iris-setosa Iris-versicolor Iris-virginica</div><div></div><div><div>Add / remove data points <input checked="" type="radio"/> Add points Iris-setosa <input type="radio"/> Remove points Remove all Open a new window</div><div>Sampling control 2 Base for sampling (r) 2 Num. locations per pixel 3 Kernel bandwidth (k)</div><div>Plotting <input checked="" type="checkbox"/> Plot training data Start</div></div></div>	<p>error = 0</p>
<p>C = 40</p> <p>gamma = 10</p>	<div><div>iris-weka.filters.unsuper... Open File</div><div>Choose LibSVM -S 0 -K 2 -D 3 -G 10.0 -R 0.0 -N 0.5 -</div><div>Class Attribute class (Nom)</div><div>Visualization Attributes X: petallength (Num) Y: petalwidth (Num)</div><div>Class color Iris-setosa Iris-versicolor Iris-virginica</div><div></div><div><div>Add / remove data points <input checked="" type="radio"/> Add points Iris-setosa <input type="radio"/> Remove points Remove all Open a new window</div><div>Sampling control 2 Base for sampling (r) 2 Num. locations per pixel 3 Kernel bandwidth (k)</div><div>Plotting <input checked="" type="checkbox"/> Plot training data Start</div></div></div>	<p>error = 6</p>

C = 50
gamma =
3

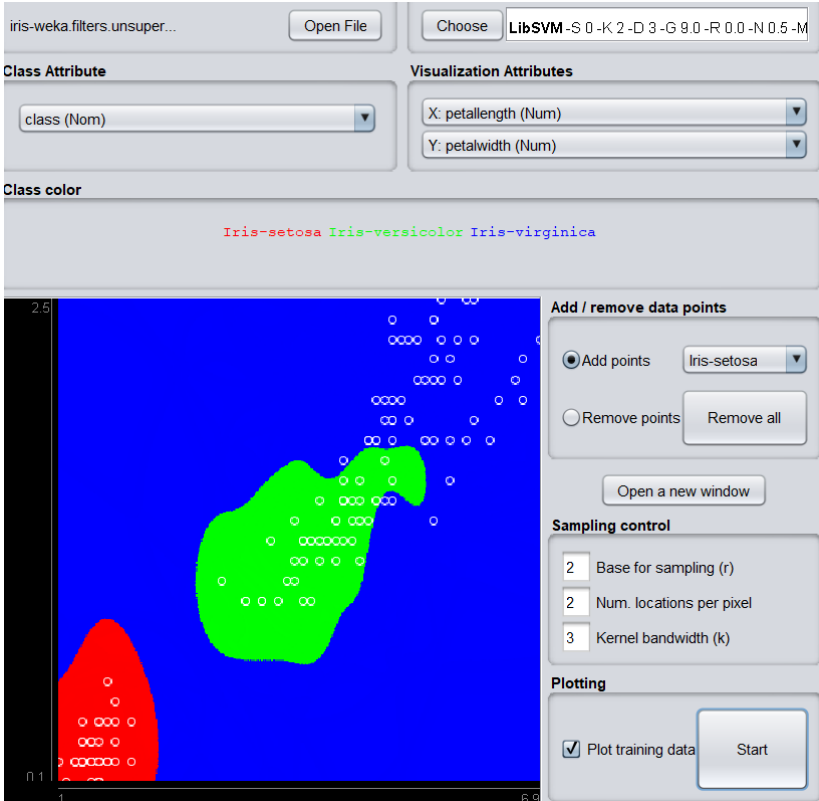
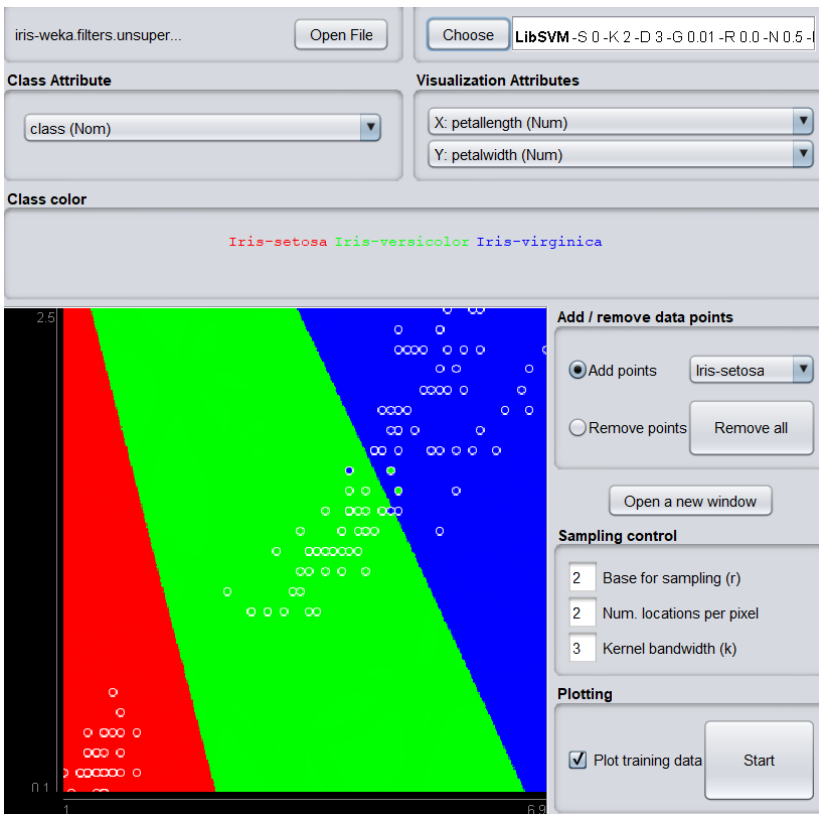


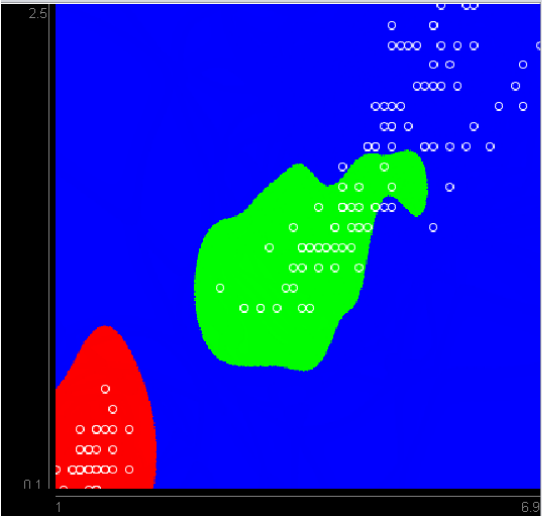
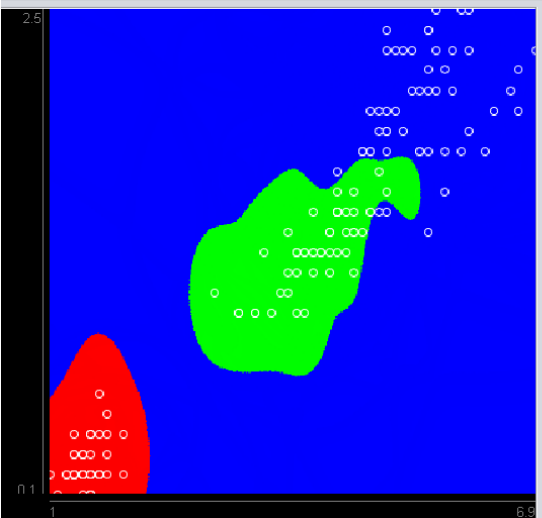
error = 2

C = 50
gamma =
6



error = 0

<p>C = 50</p> <p>gamma = 9</p>		<p>error = 1</p>
<p>C = 75</p> <p>gamma = 0,01</p>		<p>error = 4</p>

<p>C = 75</p> <p>gamma = 10</p>	<div><div>iris-weka.filters.unsuper... Open File</div><div>Choose LibSVM -S 0 -K 2 -D 3 -G 9.0 -R 0.0 -N 0.5 -M</div><div>Class Attribute class (Nom)</div><div>Visualization Attributes X: petallength (Num) Y: petalwidth (Num)</div><div>Class color Iris-setosa Iris-versicolor Iris-virginica</div><div></div><div>Add / remove data points <input checked="" type="radio"/> Add points Iris-setosa <input type="radio"/> Remove points Remove all Open a new window</div><div>Sampling control 2 Base for sampling (r) 2 Num. locations per pixel 3 Kernel bandwidth (k)</div><div>Plotting <input checked="" type="checkbox"/> Plot training data Start</div></div>	<p>error = 0</p>
<p>C = 100</p> <p>gamma = 10</p>	<div><div>iris-weka.filters.unsuper... Open File</div><div>Choose LibSVM -S 0 -K 2 -D 3 -G 10.0 -R 0.0 -N 0.5 -M</div><div>Class Attribute class (Nom)</div><div>Visualization Attributes X: petallength (Num) Y: petalwidth (Num)</div><div>Class color Iris-setosa Iris-versicolor Iris-virginica</div><div></div><div>Add / remove data points <input checked="" type="radio"/> Add points Iris-setosa <input type="radio"/> Remove points Remove all Open a new window</div><div>Sampling control 2 Base for sampling (r) 2 Num. locations per pixel 3 Kernel bandwidth (k)</div><div>Plotting <input checked="" type="checkbox"/> Plot training data Start</div></div>	<p>error = 0</p>

W tym przykładzie natomiast widzimy zdecydowanie poprawę działania. Pojawiły się bowiem przypadki, gdzie nie było żadnych błędów. Był to np. przypadek dla parametrów: $C = 17$ oraz $\gamma = 7$. LibSVM potrafi lepiej dopasować się do danych, gdyż nie ogranicza się w wyznaczaniu płaszczyzn wyłącznie poprzez proste (choć te też mogą się w nim zawierać). Tutaj, jak i poniżej najmniejszy problem stanowiła klasa, która była najbardziej oddzielona od reszty oraz była skupiona w większości tylko przy sobie. Dzięki temu klasyfikatorowi, również poprzez wizualizację, otrzymaliśmy bardzo dobre wyniki o wysokiej jakości i optymalizacji.

Omawiane dane przedstawię również za pomocą wykresu punktowego oraz wykresów hiperpłaszczyzn. Zostaną one stworzone przy pomocy środowiska R.

