

Dawid Urbaniak 204023 204023@edu.p.lodz.pl
Igor Oryński 203959 203959@edu.p.lodz.pl

Zadanie 3.: Klasyfikacja

1. Cel

Celem zadania jest rozwiązanie problemu klasyfikacji wskazanych zbiorów danych z wykorzystaniem narzędzi inteligentnej analizy danych (w tym perceptronu wielowarstwowego).

2. Wprowadzenie

Badania są przeprowadzane na dwóch ogólnodostępnych zbiorach danych:

- *Iris Data Set*,
- *Wine Data Set*.

Do klasyfikacji zbiorów wykorzystujemy następujące klasyfikatory:

- Perceptron wielowarstwowy (z poprzedniego zadania),
- SVM (support vector machine).

Zbiory danych zostały podzielone na część uczącą oraz część testową. Z każdego zbioru zostało losowo wydzielone 15 próbek danych (po 5 z każdego rodzaju) i przypisane do odpowiednich zbiorów uczących, pozostałe próbki zostały przydzielone jako zbiory testowe. Aby ułatwić operacje na danych wejściowych, zbiór *Wine Data Set* jest normalizowany do wartości z zakresu (0,1).

3. Opis implementacji

Aplikacja została zaimplementowana w języku **Python3** korzystając z bibliotek **numpy**, **SVC** oraz **gnuplot** do tworzenia wykresów.

Wszystkie stworzone metody służące do obliczeń znajdują się w przestrzeni klasy **BackPropagationNetwork**, w tym:

- **sigmoid** - metoda obliczająca sigmoidalną funkcję aktywacji,
- **trainEpoch** - metoda trenująca sieć.

Oprócz tego do najważniejszych wykorzystywanych metod możemy zaliczyć:

- **normaliseVector**- metoda normalizująca zbiór danych
- **test** - metoda testująca poprawność klasyfikacji

4. Materiały i metody

Dla każdego zbioru badana jest klasyfikacja za pomocą perceptronu oraz SVM. Badamy poprawność klasyfikacji zarówno na zbiorze testowym jak i uczącym.

W przypadku użycia perceptronu jak klasyfikatora, wykorzystujemy następujące wartości:

Dla *Iris Data Set*:

- Bias - tak
- Kolejność wzorców - losowa
- Wsp. nauki - 0.6
- Wsp. momentum - 0.0

Dla *Wine Data Set*:

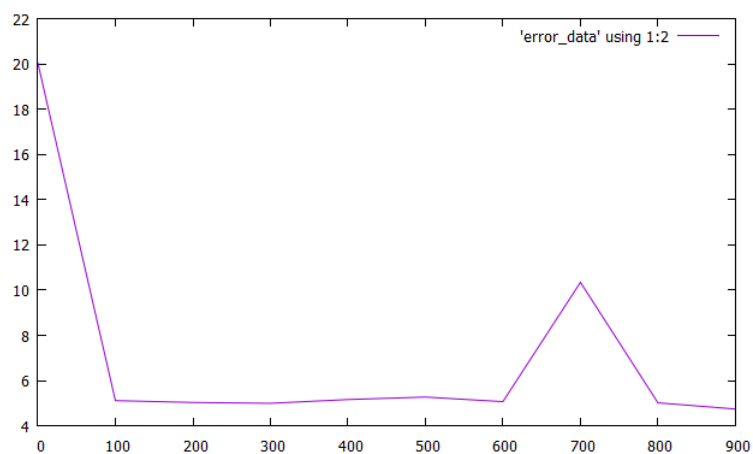
- Bias - tak
- Kolejność wzorców - stała
- Wsp. nauki - 0.2
- Wsp. momentum - 0.0

Błędy I i II rodzaju są prezentowane w macierzy pomyłek (odpowiednio nad i pod przekątną).

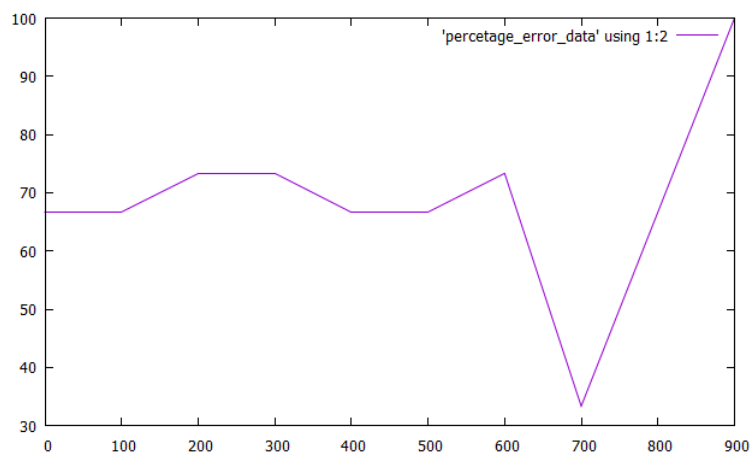
5. Wyniki

5.1. Iris Data Set

Zbiór danych uczących	iris_learn.data
Zbiór danych testowych	iris_learn.data
Klasyfikator	Perceptron
Kształt sieci	4, 10, 3



Rysunek 1. Wykres wartości błędów.



Rysunek 2. Wykres procentowej wartości zakwalifikowanych próbek.

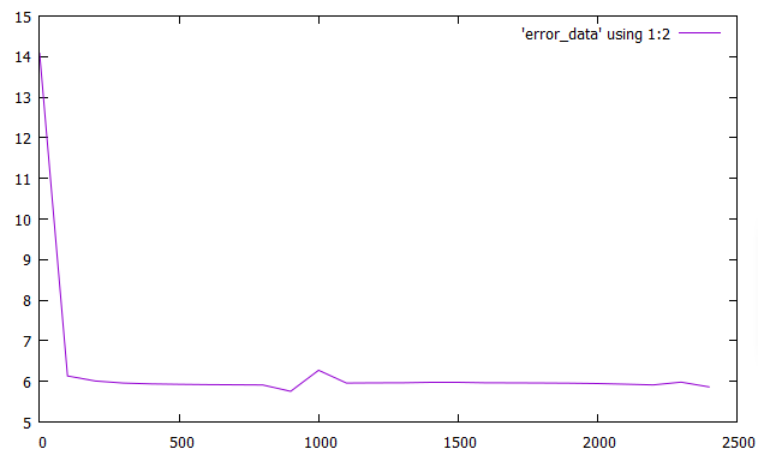
Macierz pomyłek

	I	II	III
I	5	0	0
II	0	5	0
III	0	0	5

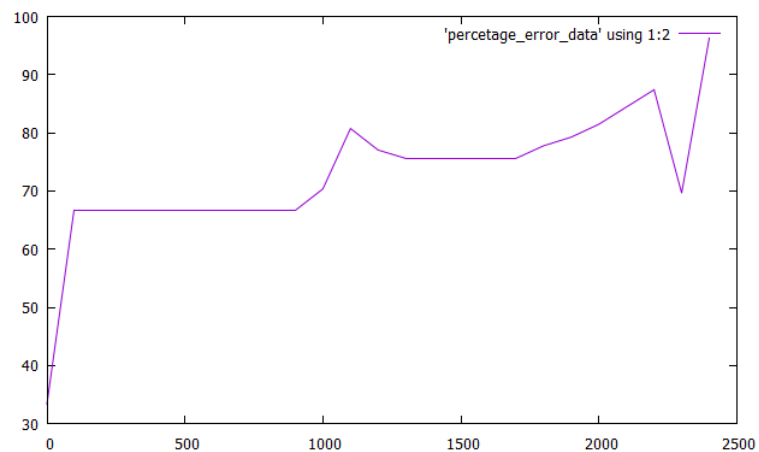
% poprawnych - 100%

Poprawnie zakwalifikowano 15 z 15 próbek.

Zbiór danych uczących	<code>iris_learn.data</code>
Zbiór danych testowych	<code>iris.data</code>
Klasyfikator	Perceptron
Kształt sieci	4, 10, 3



Rysunek 3. Wykres wartości błędu.



Rysunek 4. Wykres procentowej wartości zakwalifikowanych próbek.

Macierz pomyłek

	I	II	III
I	45	0	0
II	0	40	5
III	0	0	45

% poprawnych - 96.3%

Poprawnie zakwalifikowano 130 z 135 próbek.

Zbiór danych uczących	<code>iris_learn.data</code>
Zbiór danych testowych	<code>iris_learn.data</code>
Klasyfikator	SVM

Macierz pomyłek

	I	II	III
I	5	0	0
II	0	5	0
III	0	0	5

% poprawnych - 100%

Poprawnie zakwalifikowano 15 z 15 próbek.

Zbiór danych uczących	<code>iris_learn.data</code>
Zbiór danych testowych	<code>iris.data</code>
Klasyfikator	SVM

Macierz pomyłek

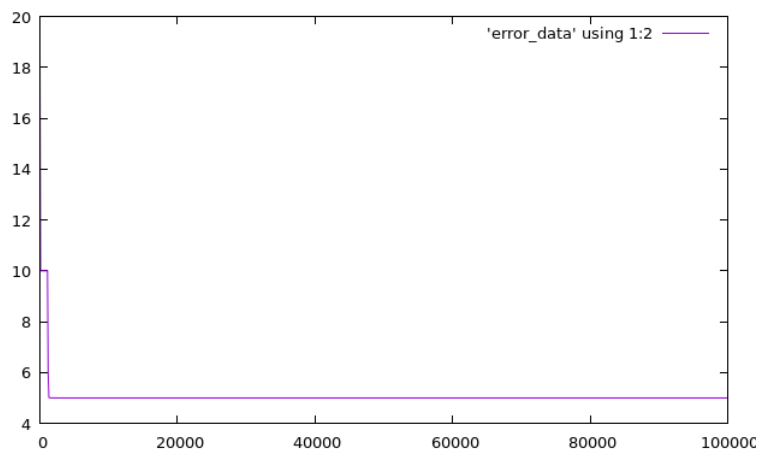
	I	II	III
I	45	0	0
II	0	42	3
III	0	1	44

% poprawnych - 97.04%

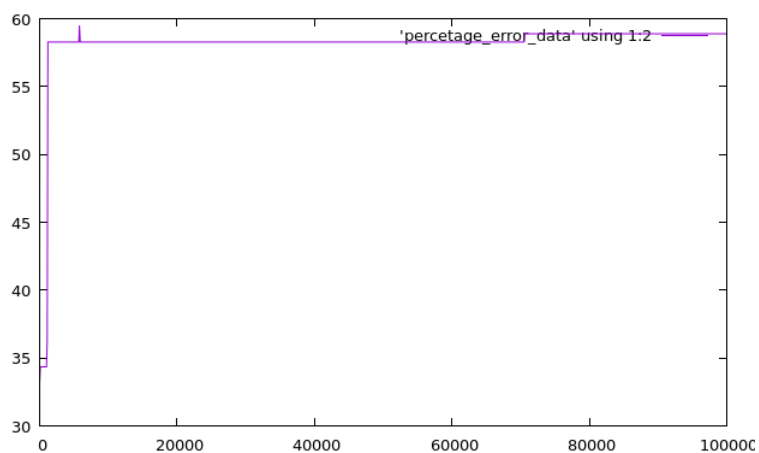
Poprawnie zakwalifikowano 131 z 135 próbek.

5.2. Wine Data Set

Zbiór danych uczących	wine_learn.data
Zbiór danych testowych	wine.data
Klasyfikator	Perceptron
Kształt sieci	4, 323, 3



Rysunek 5. Wykres wartości błędu.



Rysunek 6. Wykres procentowej wartości zakwalifikowanych próbek.

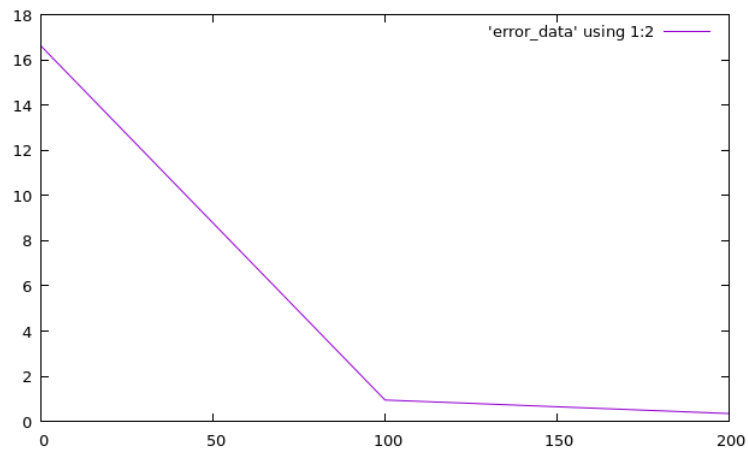
Macierz pomyłek

	I	II	III
I	54	0	0
II	50	0	16
III	1	0	42

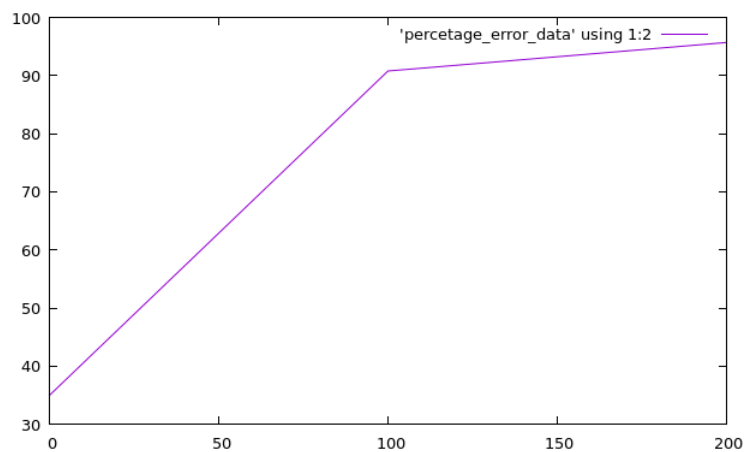
% poprawnych - 58.9%

Poprawnie zakwalifikowano 96 z 163 próbek.

Zbiór danych uczących	wine_learn.data
Zbiór danych testowych	wine.data
Klasyfikator	Perceptron
Kształt sieci	4, 97, 3



Rysunek 7. Wykres wartości błędu.



Rysunek 8. Wykres procentowej wartości zakwalifikowanych próbek.

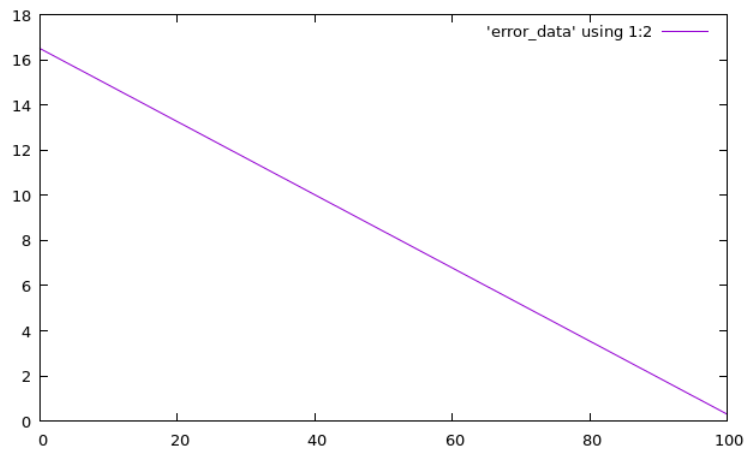
Macierz pomyłek

	I	II	III
I	51	3	0
II	0	66	0
III	1	2	39

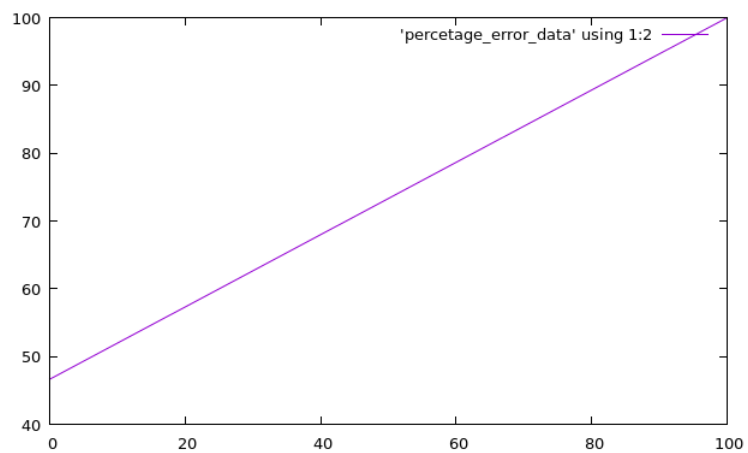
% poprawnych - 95.7%

Poprawnie zakwalifikowano 156 z 163 próbek.

Zbiór danych uczących	wine_learn.data
Zbiór danych testowych	wine_learn.data
Klasyfikator	Perceptron
Kształt sieci	4, 97, 3



Rysunek 9. Wykres wartości błędu.



Rysunek 10. Wykres procentowej wartości zakwalifikowanych próbek.

Macierz pomyłek

	I	II	III
I	5	0	0
II	0	5	0
III	0	0	5

% poprawnych - 100%

Poprawnie zakwalifikowano 15 z 15 próbek.

Zbiór danych uczących	<code>wine_learn.data</code>
Zbiór danych testowych	<code>wine_learn.data</code>
Klasyfikator	SVM

Macierz pomyłek

	I	II	III
I	0	5	0
II	0	5	0
III	0	5	0

% poprawnych - 33.33%

Poprawnie zakwalifikowano 5 z 15 próbek.

Zbiór danych uczących	<code>wine_learn.data</code>
Zbiór danych testowych	<code>wine.data</code>
Klasyfikator	SVM

Macierz pomyłek

	I	II	III
I	39	15	0
II	1	65	0
III	0	4	39

% poprawnych - 87.73%

Poprawnie zakwalifikowano 143 z 163 próbek.

6. Dyskusja

Dla zbioru *Iris Data Set* uzyskaliśmy zadowalające wyniki już przy 10 neuronach w warstwie ukrytej, uzyskując 100% poprawność przy zbiorze uczącym oraz 96.3% przy zbiorze testowym. Dla porównania, SVM uzyskał bardzo zbliżone wyniki, uzyskując skuteczność większą od perceptronu tylko o 0.74%.

W przypadku *Wine Data Set* już przy 97 neuronach uzyskaliśmy zgodność na poziomie 95.7%, czyli o 7.97% wyższą niż przy użyciu SVM. Dalsze testy na wyższej ilości neuronów nie przynosiły lepszych rezultatów, wręcz przeciwnie. Użycie 323 neuronów w warstwie ukrytej obniżyło efektywność klasyfikatora do 58.9%. Dalsze badania przynosiły podobne rezultaty, większa ilość neuronów skutkowała podobną lub mniejszą ilością prawidłowo zakwalifikowanych próbek.

Dla zbioru *Wine Data Set* musieliśmy początkowo przeprowadzić normalizację wartości do przedziału $<0,1>$ ponieważ bez niej ani perceptron ani SVM nie potrafił sklasyfikować próbek testowych ze skutecznością większą niż 50%. Po normalizacji próbek, SVM osiągnął 87% skuteczności. Nie jest to zbyt wysoki wynik lecz co warto zauważyć, nauka SVM jest zdecydowanie szybsza od perceptronu. Perceptron po znormalizowaniu próbek uzyskiwał bardzo różne wyniki w zależności od ilości neuronów w warstwie ukrytej.

7. Wnioski

- Dla zbioru *Iris Data Set* 10 neuronów zapewnia zadowalający efekt,
- 97 neuronów powoduje dokładniejszą klasyfikację zbioru *Wine Data Set* niż SVM,
- Zwiększenie liczby neuronów przynosi zmniejszenie skuteczności klasyfikacji przez perceptron,
- Normalizacja zbioru *Wine Data Set* znacząco zwiększyła poprawność klasyfikacji za pomocą SVM,
- Nauka SVM odbywa się w zdecydowanie krótszym czasie niż perceptronu wielowarstwowego.

Literatura

- [1] P. Tarasiuk [<https://ftims.edu.p.lodz.pl/>] *Zadanie klasyfikacji*