

SVM

Maszyny Wektorów Wspierających

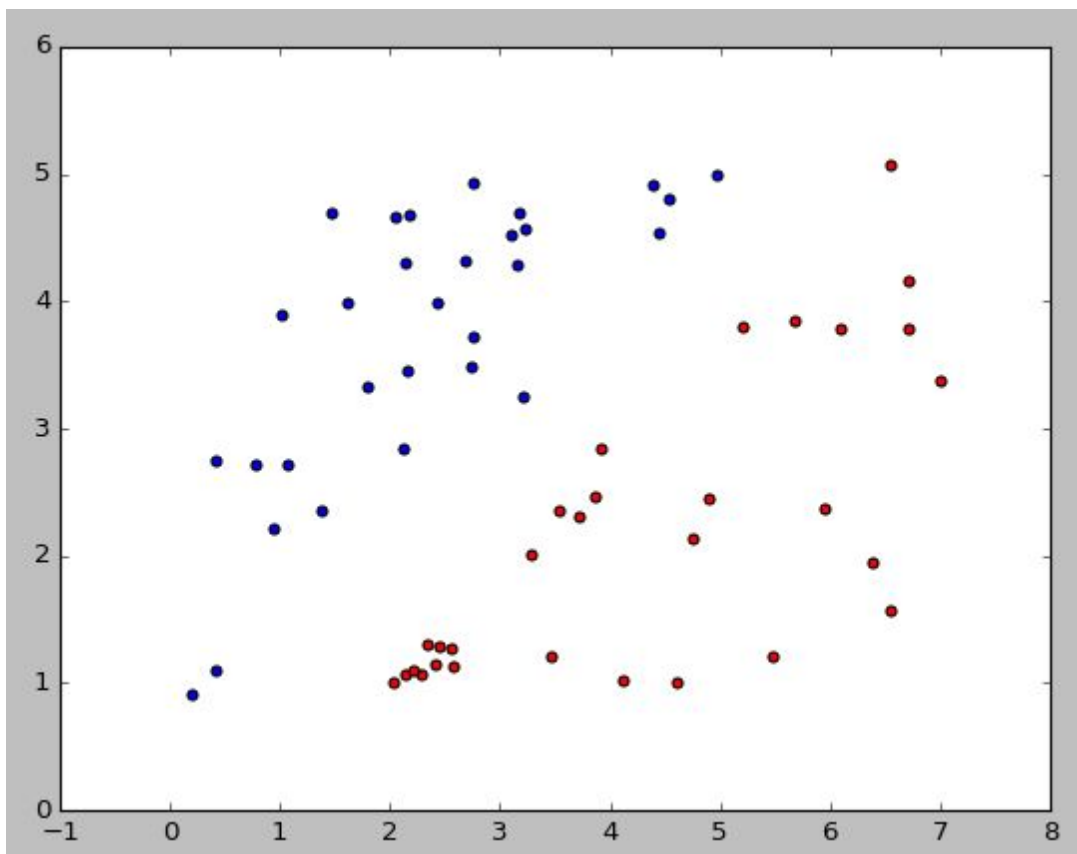
Paulina Żak

zadanie 5

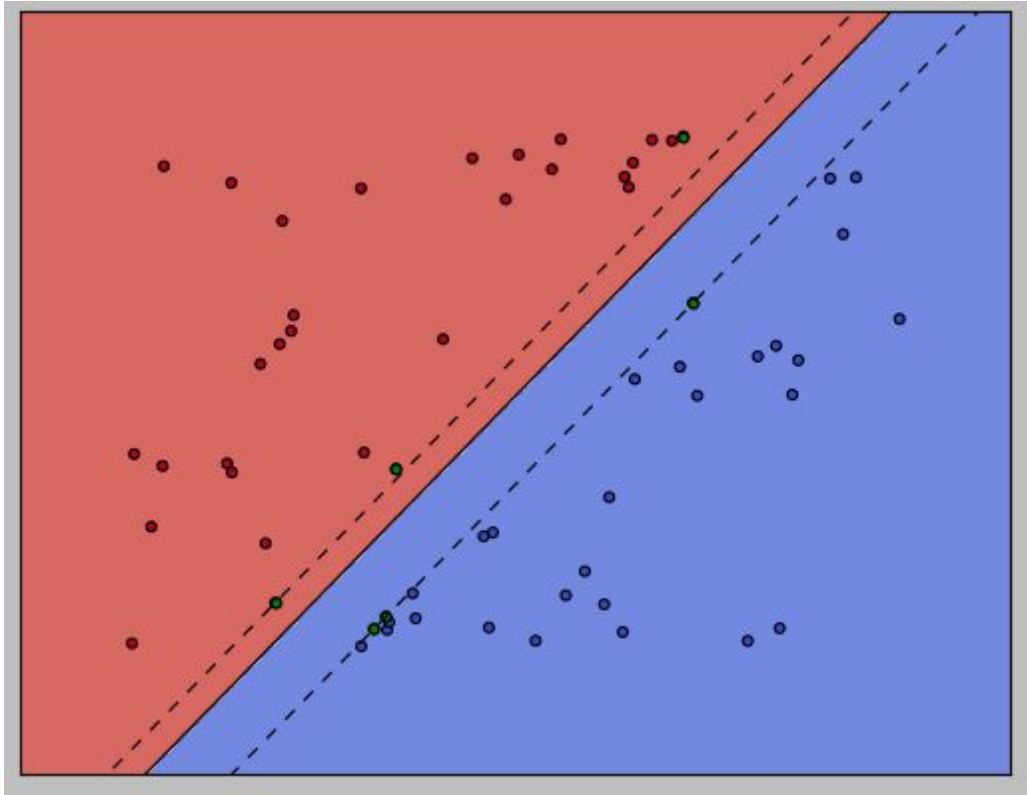
Płaszczyzna separująca o maksymalnym marginesie

Wygenerowane zostało 60 punktów:

- Niebieskie punkty wewnątrz trójkąta $[(0,0),(0,5),(5,5)]$
- Czerwone punkty wewnątrz trójkąta $[(2,6),(2,6),(2,6)]$



Hiperpłaszczyzna separująca, została wykreślona jako linia ciągła między dwoma punktami.
Linie przerywana - hiperpłaszczyzny równoległe do hiperpłaszczyzny separującej i przechodzące przez wektory wspierające
Zielony - punkty określające wektory wspierające



6. Klasyfikacja - trzy klasy

Wyniki dla wszystkich metod klasyfikacji zbioru iris okazały się zbliżone do 1.0

Całkowicie poprawnie został sklasyfikowany zbiór setosa, gdzie widać to na wykresie jako zbiór oznaczony na niebiesko. Zbiory versicolor i virginica były bardziej zbliżone do siebie i mają trochę niższe wyniki niż setosa, niemniej jednak nadal oscylujące w przedziale 0.87-1.0.

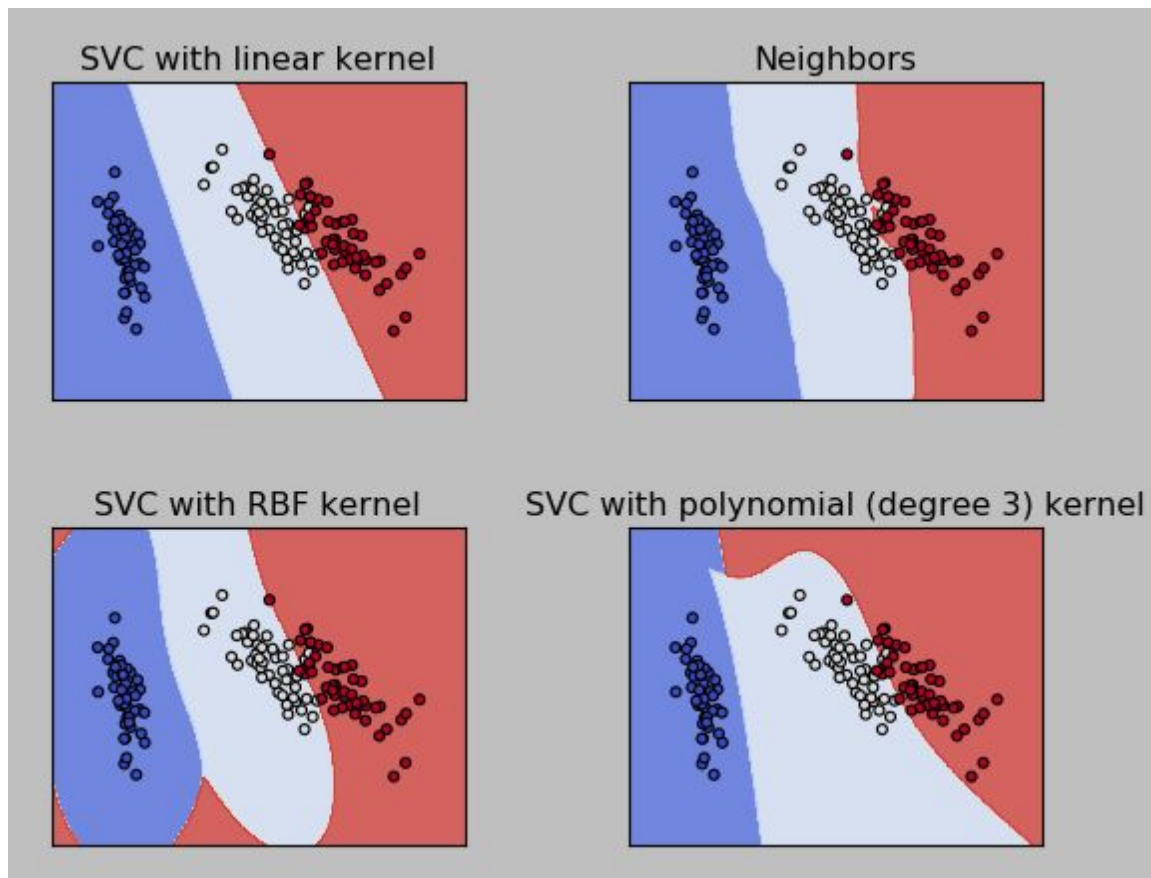
Wyniki dla zbioru testowego 0.33

| Neighbors | | | |
|------------------------|-----------|--------|----------|
| | precision | recall | f1-score |
| setosa | 1.00 | 1.00 | 1.00 |
| versicolor | 1.00 | 0.95 | 0.98 |
| virginica | 0.93 | 1.00 | 0.96 |
| avg/total | 0.98 | 0.98 | 0.98 |
| SVC with linear kernel | | | |
| | precision | recall | f1-score |
| setosa | 1.00 | 1.00 | 1.00 |
| versicolor | 1.00 | 0.95 | 0.98 |
| virginica | 0.93 | 1.00 | 0.96 |
| avg/total | 0.98 | 0.98 | 0.98 |
| SVC with RBF kernel | | | |
| | precision | recall | f1-score |
| setosa | 1.00 | 1.00 | 1.00 |
| versicolor | 1.00 | 0.91 | 0.95 |
| virginica | 0.87 | 1.00 | 0.93 |
| avg/total | 0.97 | 0.96 | 0.96 |

SVC with polynomial
(degree 3) kernel

| | precision | recall | f1-score |
|------------|-----------|--------|----------|
| setosa | 1.00 | 1.00 | 1.00 |
| versicolor | 0.95 | 0.95 | 0.95 |
| virginica | 0.92 | 0.92 | 0.92 |
| avg/total | 0.96 | 0.96 | 0.96 |

Wykres z użyciem PCA dla powyżej wymienionych metod na zbiorze iris:



7 Funkcje jądra

Wygenerowane zostały dwa nieseparowalne liniowo klastry punktów:

- Niebieskie: Pierścień o $r = 2$ i $R = 4$ stworzony z 200 punktów
- Czerwone: Pierścień o $r = 5$ i $R = 8$ stworzony z 200 punktów

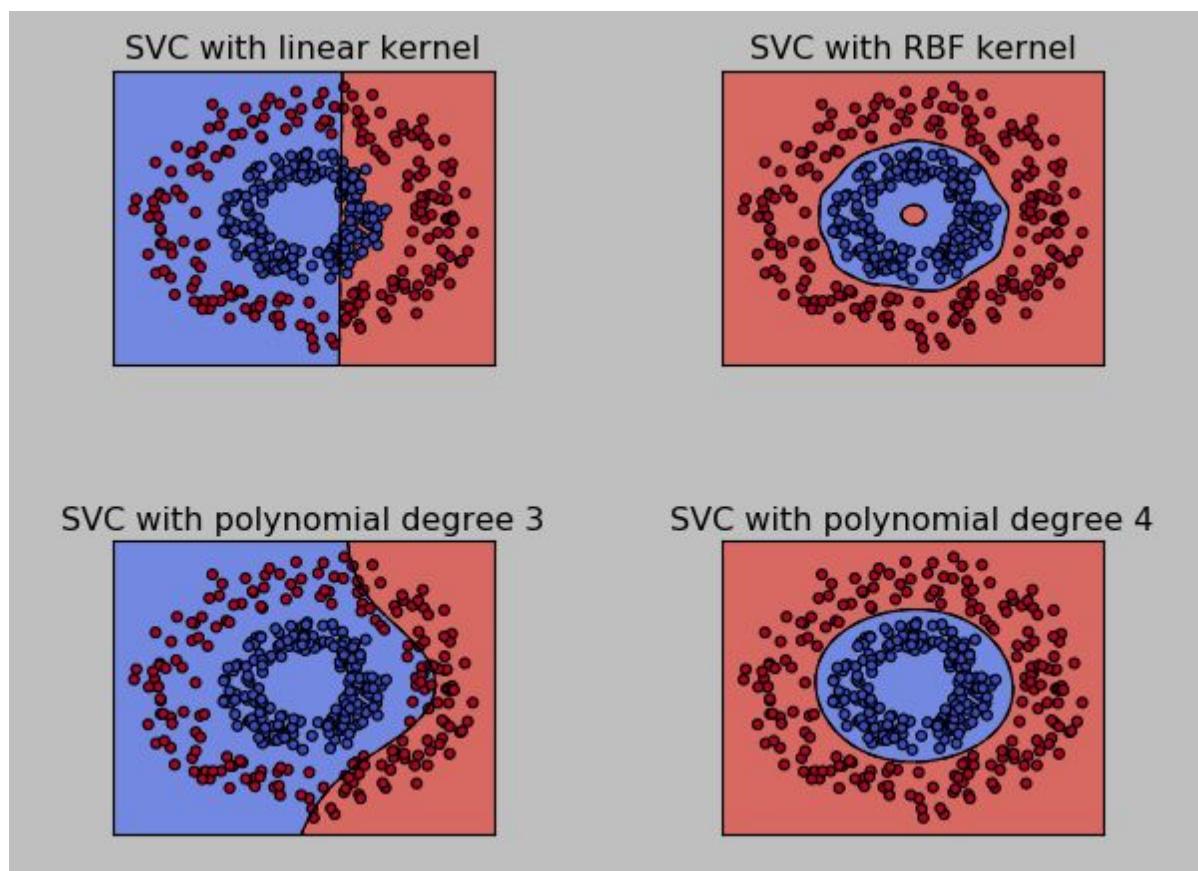
Wybrane zostały następujące funkcje kernela:

- liniowa
- rbf
- wielomianowa stopnia 4
- wielomianowa stopnia 3

Najlepiej z zadaniem oddzielenia dwóch klastrów poradził sobie 'polynomial kernel with degree 4'. Oddzielił prawidłowo oba zestawy i nie doprowadził do overfittingu.

Niemniej jednak rbf również poradziło sobie dobrze z tym zadaniem obejmując oba zestawy w dwóch oddzielnych grupach.

Natomiast kernel liniowy jak i wielomianowy stopnia 3 nie były w stanie poradzić sobie z tak skomplikowanym zbiorem.



8 Rozpoznawanie twarzy z zastosowaniem projekcji eigenfaces



W eigenfaces wzmocniane są takie elementy twarzy jak nos, usta i włosy.

Number of samples

1288

Number of features

62500

Number of classes

7

Image shapes

250 250

W wygenerowanej tablicy pomyłek (confusion matrix)

Widzimy, że program nigdy nie zweryfikował obrazu jako zdjęcia Hugo_Chavez'a.

Dlatego też jego precision recall jak i f1-score są zerowe.

Dodatkowo, jeżeli było to zdjęcie George W Busha'a to w 163/164 przypadkach zostało określone poprawnie, jednocześnie wiele zdjęć innych osób także zostało określone jako zdjęcia Busha.

Co więcej jeżeli program ocenił zdjęcie jako zdjęcie Ariela Sharona bądź Donalda Rumsfelda to zawsze miał rację, dzięki temu osiągnął precyzję na tych zdjęciach równą 1.00.

Oprócz Hugo_Chavez'a, precyzja była wysoka (w przedziale 0.63-1.0), a recall wahał się pomiędzy 0.28-0.99, natomiast f1-score 0.43-0.77

Tablica pomyłek

| | Ariel Sharon | Colin Powell | Donald Rumsfeld | George W Bush | Gerhard Schroeder | Hugo Chavez | Tony Blair |
|-------------------|--------------|--------------|-----------------|---------------|-------------------|-------------|------------|
| Ariel Sharon | 5 | 2 | 0 | 11 | 0 | 0 | 0 |
| Colin Powell | 0 | 55 | 0 | 17 | 0 | 0 | 0 |
| Donald Rumsfeld | 0 | 4 | 14 | 25 | 2 | 0 | 0 |
| George W Bush | 0 | 0 | 0 | 163 | 0 | 0 | 1 |
| Gerhard Schroeder | 0 | 0 | 0 | 15 | 10 | 0 | 5 |
| Hugo Chavez | 0 | 0 | 0 | 14 | 2 | 0 | 2 |
| Tony Blair | 0 | 2 | 0 | 14 | 0 | 0 | 24 |

Podział zbioru treningowego/testowego -0.7/0.3

| | precision | recall | f1-score |
|-------------------|-----------|--------|----------|
| Ariel_Sharon | 1.00 | 0.28 | 0.43 |
| Colin_Powell | 0.87 | 0.76 | 0.81 |
| Donald_Rumsfeld | 1.00 | 0.31 | 0.47 |
| George_W_Bush | 0.63 | 0.99 | 0.77 |
| Gerhard_Schroeder | 0.71 | 0.33 | 0.45 |
| Hugo_Chavez | 0.00 | 0.00 | 0.00 |
| Tony_Blair | 0.75 | 0.60 | 0.67 |
| avg / total | 0.72 | 0.70 | 0.66 |

Wyniki pogarszają się ze zmianą wielkości zbioru treningowego

Podział zbioru treningowego/testowego -0.9/0.1

| | precision | recall | f1-score |
|-------------|-----------|--------|----------|
| avg / total | 0.69 | 0.68 | 0.64 |

Podział zbioru treningowego/testowego -0.5/0.5

| | precision | recall | f1-score |
|-------------|-----------|--------|----------|
| avg / total | 0.69 | 0.61 | 0.54 |