Sieci Neuronowe – laboratorium  
  
Ćwiczenie nr. 1  
  
Kajetan Pynka, 254495

Spis treści

[Perceptron prosty 3](#_Toc116820200)

[Eksperyment 1 3](#_Toc116820201)

[Eksperyment 2 4](#_Toc116820202)

[Eksperyment 3 5](#_Toc116820203)

[Eksperyment 4 6](#_Toc116820204)

[Adaline 7](#_Toc116820205)

[Eksperyment 1 7](#_Toc116820206)

[Eksperyment 2 8](#_Toc116820207)

[Eksperyment 3 9](#_Toc116820208)

[Eksperyment 4 10](#_Toc116820209)

# Wstęp

Rozmiar danych = 100

Podział na zbiór treningowy / testowy: 80/20

Domyślny wsp. uczenia = 0.01

Domyślna waga = 0.001

# Perceptron prosty

## Eksperyment 1

Ogólnie gdy theta rośnie do +inf to potrzeba więcej epok na wyuczenie modelu, zależy to bezpośrednio od współczynnika uczenia: małe kroki → dłużej zajmuje osiągnięcie wag przekraczających próg. Dla wartości bardzo bliskch zera potrzeba mieć adekwatnie mały współczynnik uczenia. Dla ujemnych theta oraz funkcji aktywacji unipolarnej model nigdy się nie wyuczy (zawsze predykcja będzie dobra → niezerowa delta nigdy się nie skoryguje).

|  |  |  |
| --- | --- | --- |
| **Próg odcięcia (theta)** | **Średnia liczba epok potrzebna do wyuczenia** | **Średnie wagi [x1, x2]** |
| 0.05 | 3.0 | [0.0317, 0.0301] |
| 0.10 | 4.0 | [0.0594, 0.0598] |
| 0.15 | 4.0 | [0.0886, 0.0897] |
| 0.20 | 5.0 | [0.1172, 0.1194] |
| 0.25 | 6.0 | [0.1452, 0.1510] |
| 0.30 | 6.0 | [0.1678, 0.1721] |
| 0.35 | 6.3 | [0.1959, 0.2028] |
| 0.40 | 7.6 | [0.2239, 0.2312] |
| 0.45 | 8.0 | [0.2540, 0.2611] |
| 0.50 | 9.0 | [0.2818, 0.2932] |
| 0.80 | 12.2 | [0.4482, 0.4640] |
| 2.50 | 34.0 | [1.3804, 1.4340] |
| 100.00 | 1270.3 | [54.941, 57.105] |

## Eksperyment 2

Jeśli wylosuje się idealne wagi to w zasadzie model może być od razu wyuczony, w związku z tym im bliżej optimum zostaną zainicjalizowane tym mniej epok będzie wymaganych do wyuczenia modelu. Najczęściej trudne będzie oszacowanie potencjalnie optymalnych wag, dlatego należy zwężać zakres wag „o tyle, ile się da”.

|  |  |  |
| --- | --- | --- |
| **Zakres wag** | **Średnia liczba epok potrzebna do wyuczenia** | **Średnie wagi [x1, x2, bias]** |
| -1.0 do 1.0 | 10.0 | [0.2870, 0.3125, -0.4560] |
| -0.9 do 0.9 | 10.6 | [0.1581, 0.1280, -0.2335] |
| -0.8 do 0.8 | 10.7 | [0.1141, 0.2055, -0.2632] |
| -0.7 do 0.7 | 7.8 | [0.1168, 0.1375, -0.1883] |
| -0.6 do 0.6 | 8.3 | [0.1293, 0.1238, -0.2028] |
| -0.5 do 0.5 | 8.1 | [0.1233, 0.1278, -0.2168] |
| -0.4 do 0.4 | 6.1 | [0.0781, 0.0952, -0.1404] |
| -0.3 do 0.3 | 6.0 | [0.0781, 0.0795, -0.1223] |
| -0.2 do 0.2 | 5.2 | [0.0471, 0.0836, -0.1113] |
| -0.1 do 0.1 | 4.4 | [0.0279, 0.0284, -0.0455] |

## Eksperyment 3

Nieważne jak duży albo jak mały współczynnik uczenia damy, model w podobnej liczbie epok będzie w stanie się wyuczyć **(przy założeniu wykorzystania biasu!).** Jeśli manualnie ustalamy próg theta to bardzo istotne jest dobranie współczynnika uczenia, ponieważ za duży może skutkować zbyt dużymi zmianami wag (co uniemożliwi modelowi osiągnięcie optymalnego rozwiązania).

|  |  |  |
| --- | --- | --- |
| **Wartość współczynnika uczenia** | **Średnia liczba epok potrzebna do wyuczenia** | **Średnie wagi [x1, x2, bias]** |
| 5000.0000 | 5.0 | [10261.7, 11323.4, -15000.0] |
| 1.0000 | 4.9 | [1.9861, 2.2482, -3.0000] |
| 0.9000 | 4.6 | [1.7569, 1.7883, -2.4303] |
| 0.8000 | 4.8 | [1.6106, 1.6991, -2.4001] |
| 0.7000 | 4.9 | [1.4389, 1.5441, -2.1000] |
| 0.6000 | 5.0 | [1.2312, 1.3587, -1.7998] |
| 0.5000 | 4.8 | [1.0263, 1.0691, -1.5000] |
| 0.4000 | 5.2 | [0.8595, 0.8770, -1.2399] |
| 0.3000 | 5.0 | [0.6061, 0.6781, -0.8997] |
| 0.2000 | 4.9 | [0.4053, 0.4526, -0.5999] |
| 0.1000 | 5.0 | [0.1988, 0.2256, -0.2996] |
| 0.0001 | 3.4 | [0.0003, 0.0003, -0.0005] |

## Eksperyment 4

Lepiej sprawdza się bipolarna funkcja aktywacji, wynika to prawdopodobnie z faktu, że korekcje wag są silniejsze (zbiór wartości różnicy poprawnej odpowiedzi oraz predykcji wynosi {-2, 0, 2} dla funkcji bipolarnej natomiast {-1, 0, 1} dla funkcji unipolarnej).

|  |  |  |
| --- | --- | --- |
| **Rodzaj funkcji aktywacji** | **Średnia liczba epok potrzebna do wyuczenia** | **Średnie wagi [x1,x2,bias]** |
| unipolarna | 4.2 | [0.0199, 0.0189, -0.0289] |
| bipolarna | 3.0 | [0.0217, 0.0190, -0.0198] |

# ADALINE

## Eksperyment 1

|  |  |  |
| --- | --- | --- |
| **Zakres wag** | **Średnia liczba epok potrzebna do wyuczenia** | **Średnie wagi [x1,x2,bias]** |
| -1.0 do 1.0 | 3.5 | [0.4230, 0.4320, -0.5440] |
| -0.9 do 0.9 | 3.0 | [0.3735, 0.4091, -0.5563] |
| -0.8 do 0.8 | 3.3 | [0.3887, 0.4096, -0.5459] |
| -0.7 do 0.7 | 3.2 | [0.3784, 0.4543, -0.5160] |
| -0.6 do 0.6 | 3.3 | [0.3583, 0.4098, -0.4789] |
| -0.5 do 0.5 | 3.0 | [0.3739, 0.4041, -0.4546] |
| -0.4 do 0.4 | 2.9 | [0.3874, 0.3875, -0.4845] |
| -0.3 do 0.3 | 3.0 | [0.3872, 0.4156, -0.4732] |
| -0.2 do 0.2 | 3.0 | [0.3702, 0.3948, -0.4883] |
| -0.1 do 0.1 | 3.0 | [0.3718, 0.4064, -0.4729] |

## Eksperyment 2

|  |  |  |
| --- | --- | --- |
| **Współczynnik uczenia** | **Średnia liczba epok potrzebna do wyuczenia** | **Średnie wagi [x1,x2,bias]** |
| 0.50000 | NULL | NULL |
| 0.30000 | 2.0 | [0.9637, 0.9342, -1.0302] |
| 0.20000 | 2.0 | [0.8721, 0.8847, -0.9160] |
| 0.10000 | 2.0 | [0.7105, 0.7625, -0.7555] |
| 0.05000 | 2.0 | [0.5743, 0.6346, -0.6542] |
| 0.02500 | 3.0 | [0.4992, 0.5462, -0.5961] |
| 0.01000 | 3.0 | [0.3696, 0.4069, -0.4785] |
| 0.00100 | 11.0 | [0.2387, 0.2659, -0.3299] |
| 0.00010 | 94.0 | [0.2257, 0.2516, -0.3143] |
| 0.00001 | 922.2 | [0.2243, 0.2499, -0.3122] |

## Eksperyment 3

|  |  |  |
| --- | --- | --- |
| **Dopuszczalny błąd** | **Średnia liczba epok potrzebna do wyuczenia** | **Średnie wagi [x1,x2,bias]** |
| 1.0000 | 2.0 | [0.2542, 0.2861, -0.3422] |
| 0.7500 | 2.0 | [0.2543, 0.2862, -0.3420] |
| 0.5000 | 3.0 | [0.3697, 0.4069, -0.4785] |
| 0.4500 | 3.0 | [0.3696, 0.4068, -0.4785] |
| 0.4000 | 3.0 | [0.3696, 0.4068, -0.4785] |
| 0.3500 | 3.0 | [0.3696, 0.4068, -0.4785] |
| 0.3000 | 3.1 | [0.3749, 0.4120, -0.4839] |
| 0.2500 | 5.0 | [0.4461, 0.4809, -0.5533] |
| 0.2485 | 5.0 | [0.4461, 0.4809, -0.5533] |
| 0.2000 | NULL | NULL |

## Eksperyment 4

Wsp. uczenia dla perceptronu może być giga duży bo i tak pobudzenie jest rzutowane funkcją aktywacji. Dla ADALINE jak się da za duży wsp. uczenia to wagi mogą uciekać do +- nieskończoności bo za duża delta będzie wychodzić.