**Wprowadzenie do Sztucznej Inteligencji**

**Laboratorium 4**

SVM oraz drzewa decyzyjne

Piotr Niedziałek

304474

1. **Wstęp**

Przedmiotem laboratorium były dwa algorytmy uczenia maszynowego – SVM oraz drzewa decyzyjne. Są one szeroko stosowane do zadania klasyfikacji. SVM najlepiej spełnia swoją rolę dla zadań liniowo separowalnych, jednakże przy użyciu funkcji jądra można go również skutecznie stosować dla pozostałego typu zadań. Drzewa decyzyjne najczęściej stosowane są dla danych nieliniowych i gdy istotne jest określenie ,która cecha wpłynęła na decyzję.

1. **Zadanie**

W laboratorium należało przeprowadzić serię eksperymentów, ucząc oba algorytmy na zbiorze danych z klasyfikacją irysów, ustawiając różne parametry. Dla SVM zmieniano siłę regularyzacji, funkcję jądra oraz maksymalną ilość iteracji. Dla drzewa decyzyjnego manipulowano kryterium oceny, techniką podziału węzła oraz maksymalną głębokością drzewa. Jako miary jakości klasyfikacji użyto accuracy, precision, recall oraz F1.

1. **Wyniki**
   1. **SVM**

Siła regularyzacji

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **C** | **ACCURACY** | **PRECISSION** | **RECALL** | **F1** |
| 0.001 | 0.913 0.0542 | 0.92  0.0516 | 0.91330.0541 | 0.9123 0.0552 |
| 0.01 | 0.920 0.0400 | 0.928 0.0373 | 0.92 0.04 | 0.9191 0.04 |
| 0.1 | 0.973 0.0133 | 0.976 0.0121 | 0.973 0.013 | 0.9732 0.0133 |
| 1 | 0.980  0.0163 | 0.982  0.0148 | 0.980  0.0163 | 0.979  0.0164 |
| 10 | 0.973  0.0389 | 0.979  0.0301 | 0.973  0.0389 | 0.9728  0.0397 |
| 100 | 0.967  0.0422 | 0.971  0.0353 | 0.967  0.0422 | 0.966  0.0428 |
| 1000 | 0.966  0.0422 | 0.971  0.0353 | 0.967  0.0422 | 0.966  0.0428 |

Funkcja jądra

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KERNEL** | **ACCURACY** | **PRECISSION** | **RECALL** | **F1** |
| linear | 0.98 0.0163 | 0.9818  0.0148 | 0.9800.0163 | 0.9790.01637 |
| poly | 0.98 0.0163 | 0.9818  0.0148 | 0.9800.0163 | 0.9790.01637 |
| Rbf | 0.966 0.0211 | 0.968 0.0211 | 0.966 0.0210 | 0.967 0.0210 |
| sigmoid | 0.067  0.0596 | 0.0318  0.0247 | 0.066  0.059 | 0.0428  0.0349 |

Ilość iteracji

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ITERACJE** | **ACCURACY** | **PRECISSION** | **RECALL** | **F1** |
| 100 | 0.980 0.0163 | 0.9818  0.0148 | 0.9800.0163 | 0.9790.01637 |
| 1000 | 0.980 0.0163 | 0.9818  0.0148 | 0.9800.0163 | 0.9790.01637 |
| 10000 | 0.980 0.0163 | 0.9818  0.0148 | 0.9800.0163 | 0.9790.01637 |
| 100000 | 0.980 0.0163 | 0.9818  0.0148 | 0.9800.0163 | 0.9790.01637 |

* 1. **Drzewo decyzyjne**

Kryterium oceny

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KRYTERIUM** | **ACCURACY** | **PRECISSION** | **RECALL** | **F1** |
| entropy | 0.953 0.0340 | 0.955  0.0337 | 0.9530.0334 | 0.9530.0340 |
| gini | 0.958 0.0331 | 0.959  0.0327 | 0.9570.033 | 0.9570.0331 |
| log\_loss | 0.953 0.0340 | 0.955  0.0337 | 0.9530.0334 | 0.9530.0340 |

Technika podziału węzłą

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PODZIAŁ** | **ACCURACY** | **PRECISSION** | **RECALL** | **F1** |
| best | 0.953 0.0340 | 0.9577  0.0331 | 0.95770.0331 | 0.95770.0331 |
| random | 0.958 0.0331 | 0.9511  0.0269 | 0.9510.0269 | 0.9510.0274 |

Maksymalna głębokość drzewa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GŁĘBOKOŚĆ** | **ACCURACY** | **PRECISSION** | **RECALL** | **F1** |
| 10 | 0.957 0.0331 | 0.95986  0.0327 | 0.9570.0331 | 0.9570.0332 |
| 100 | 0.957 0.0331 | 0.95986  0.0327 | 0.9570.0331 | 0.9570.0332 |
| 1000 | 0.957 0.0331 | 0.95986  0.0327 | 0.9570.0331 | 0.9570.0332 |
| 10000 | 0.957 0.0331 | 0.95986  0.0327 | 0.9570.0331 | 0.9570.0332 |