

Maciej Groszyk 289761

Zadanie 7 - wsi: Klasyfikator naiwny bayesa

Treść zadania:

Zaimplementuj naiwny klasyfikator Bayesa oraz zbadaj działanie algorytmu w zastosowaniu do zbioru danych Iris Data Set. Pamiętaj, aby podzielić zbiór danych na zbiór trenujący oraz uczący.

Opis rozwiązania:

Naiwny klasyfikator bayesa jest to prosty klasyfikator probabilistyczny., który służy do klasyfikacji obiektów na podstawie cech do odpowiedniej klasy. Trenowanie modelu polega na analizie zbioru danych uczących w których podane są prawidłowe klasy do zbioru odpowiednich cech. W tym celu należy rozdzielić zbiór danych trenujących na podgrupy w zależności od występujących klas obiektów. Należy obliczyć prawdopodobieństwo wystąpienia każdej z klas w zbiorze, a następnie na podzielonych podzbiorach należy wyliczyć średnią oraz wariancję dla danych cech. Po obliczeniu tych wartości, algorytm jest w stanie na podstawie twierdzenia bayesa obliczyć prawdopodobieństwo z gęstości rozkładu normalnego o obliczonych parametrach.

Wynik prezentują się następująco:

Dla IRIS Dataset. Z podziałem 80/20 na grupy trenujące i testujące. Algorytm działa z prawie 97% skutecznością. Na 30 wektorów wejściowych, dobrze sklasyfikował 29 z nich.

```
mac@t812:~/Desktop/WSI/bayes$ /bin/python3 /home/mac/Desktop/WSI/bayes/bayes.py
Accuracy: 96.66666666666667 % - 29/30
mac@t812:~/Desktop/WSI/bayes$
```

W momencie podziału w stosunku 70/30, skuteczność jest równa około 91%. na 45 rekordów dobrze sklasyfikował 41 z nich.

```
mac@t812:~/Desktop/WSI/bayes$ /bin/python3 /home/mac/Desktop/WSI/bayes/bayes.py
Accuracy: 91.11111111111111 % - 41/45
mac@t812:~/Desktop/WSI/bayes$
```

Dla podziału 60/40. Skuteczność jest na poziomie 93.3%.

```
mac@t812:~/Desktop/WSI/bayes$ /bin/python3 /home/mac/Desktop/WSI/bayes/bayes.py
Accuracy: 93.33333333333333 % - 56/60
mac@t812:~/Desktop/WSI/bayes$
```

Wnioski:

Algorytm ten jest bardzo skuteczny. Pozwala z dużą dokładnością sklasyfikować dane wejściowe. Niezależnie od stosunku podziału danych wejściowych.