

DRZEWA DECYZYJNE

LAB 2 – Redukcja wymiarowości

Zadanie 1. (2 pkt.)

Wczytaj zbiór danych pochodzący z biblioteki sklearn za pomocą komendy:

```
from sklearn.datasets import load_breast_cancer
print(load_breast_cancer().data)
print(load_breast_cancer().target)
```

Zapoznaj się z informacjami dotyczącymi wczytanych danych znajdującymi się w dokumentacji sklearn.datasets.

Jako wynik działania programu wypisz następujące informacje:

1. Wymiar wczytanych danych
2. Ilość wartości unikatowych w wektorze target
3. Zaproponuj usunięcie tych kolumn ze zbioru danych, które Twoim zdaniem dostarczają najmniej informacji.
4. Zapisz w pliku dataset_cut.csv wypadkowy zbiór danych jak poniżej:
ColName1; ColName2; ...; TARGET
5. Skomentuj otrzymane rezultaty.

Zadanie 2. (2 pkt.)

Wykonaj redukcję wymiarowości za pomocą algorytmu PCA – analiza głównych składowych.

1. Wprowadź zbiór danych do algorytmu PCA
2. Wykonaj redukcję wymiarowości do 5 wymiarów wypadkowego zbioru danych
3. Zapisz w pliku dataset_pca_5.csv wypadkowy zbiór danych jak poniżej:
COMP1; COMP2; COMP3; COMP4; COMP5; TARGET
4. Oblicz wariancję zbioru danych po redukcji wymiarowości
5. Oblicz wartość wariancji wyjaśnionej dla wygenerowanych składowych (*ang. explained variance ratio*)
6. Skomentuj otrzymane rezultaty.

Zadanie 3. (3 pkt.)

Wykonaj redukcję wymiarowości za pomocą algorytmu PCA – analiza głównych składowych.

1. Uzasadnij optymalny wymiar zbioru danych zakładając, że w zbiorze wypadkowym suma składowych będzie stanowiła minimum 90% wariancji.
2. Wykonaj redukcję wymiarowości za pomocą algorytmu PCA.
3. Zapisz w pliku `dataset_pca_n.csv` wypadkowy zbiór danych jak poniżej:
COMP1; COMP2; COMP3; COMP4; ...; COMPN; TARGET
4. Oblicz wartość wariancji wyjaśnionej dla wygenerowanych składowych (*ang. explained variance ratio*)
5. Skomentuj otrzymane rezultaty.

Zadanie 4. (3 pkt.)

Wykonaj redukcję wymiarowości na podstawie dowolnego innego algorytmu a wynik zapisz w pliku `dataset_algorithm.csv`. Skomentuj otrzymane rezultaty.