

Rozpoznawanie obrazów - laboratorium

Zadanie 1

Urszula Libal

kwiecień 2025

Rozwiązania zadań proszę przygotować w Jupyter Notebook i wgrać na eportalu.
Wyniki zaprezentują Państwo podczas laboratorium.
Do zdobycia w sumie 55 punktów.

1 Klasyfikator bayesowski

1.1 Analiza teoretyczna ryzyka bayesowskiego (10 p.)

Wylicz teoretyczne ryzyko błędnej klasyfikacji algorytmu bayesowskiego (ryzyko bayesowskie).
Wylicz prawdopodobieństwo poprawnej klasyfikacji dla algorytmu bayesowskiego.
Rozkłady prawdopodobieństwa w klasach są następujące:

KLASA 1: prawdopodobieństwo a priori $p_1 = 0.5$, funkcja gęstości to

$$f_1(x) = \begin{cases} -2x + 2 & \text{dla } x \in [0, 1] \\ 0 & \text{poza przedziałem} \end{cases} \quad (1)$$

KLASA 2: prawdopodobieństwo a priori $p_2 = 0.5$, funkcja gęstości to

$$f_2(x) = \begin{cases} 0.5 & \text{dla } x \in [0.75, 2.75] \\ 0 & \text{poza przedziałem} \end{cases} \quad (2)$$

UWAGA: W Jupyter Notebook jest opcja użycia wzorów pisanych poleceniami Latex-owymi.
Proszę rozpisać obliczenia wzorami.

1.2 Empiryczna wersja algorytmu bayesowskiego (10 p.)

Dla danych wylosowanych z podanych rozkładów prawdopodobieństwa w klasach (dane w pliku "samples.csv"), przeprowadź eksperyment numeryczny polegający na klasyfikacji empiryczną wersją algorytmu bayesowskiego oraz porównaniu dokładności klasyfikacji z wynikiem teoretycznym.

1.3 Porównanie z innymi klasyfikatorami (5 p.)

Dodatkowo proszę przeprowadzić serię eksperymentów polegających na klasyfikacji danych różnymi klasyfikatorami. Porównaj z klasyfikatorem bayesowskim teoretycznym i empirycznym.
Czy klasyfikator bayesowski okazał się najlepszy?

2 Zbiory danych 'Iris', 'Digits', 'Wine', 'Breast cancer'

Dla wybranych **dwóch** zbiorów danych przeprowadź:

2.1 Wstępna analiza danych (10 p.)

Analiza danych (wizualizacja, separowalność klas, korelacja cech, rozkłady poszczególnych cech)

2.2 Klasyfikacja (10 p.)

Klasyfikacja różnymi klasyfikatorami (po podzieleniu na dane treningowe i testowe - dla każdego klasyfikatora można przeprowadzić kilka klasyfikacji zmieniając parametry). Porównanie dokładności klasyfikacji dla różnych klasyfikatorów - proszę przedstawić w tabeli.

2.3 Analiza wyników klasyfikacji (10 p.)

Analiza wyników klasyfikacji z uwzględnieniem właściwości danych otrzymanych ze wstępnej analizy danych (spełnienie założeń metody, wizualizacja granic między obszarami decyzyjnymi, wizualna separowalność klas a macierz pomyłek).