

1. Elementarna wizualizacja danych / podstawowe statystyki (analiza opisowa)
  - średnie, odchylenia standardowe, macierz korelacji
  - histogram, wykres pudełkowy
  - wykresy rozrzutu (scatterplots)Zinterpretuj wyniki.
2. Dane symulowane. Wygeneruj próby z np. dwuwymiarowego rozkładu normalnego dla różnych parametrów średniej, wariancji, korelacji. Porównaj statystyki próbkowe z prawdziwymi parametrami. Wyciągnij wnioski.
3. Analiza danych pod kątem klasyfikacji.
  - przeprowadź analizy jak w zad. 1
  - które zmienne / zestawy zmiennych charakteryzują się lepszymi zdolnościami dyskryminacyjnymi?
  - przetestuj wybrane techniki klasyfikacji na różnych podzbiorach cech (czy jest optymalny podzbiór, czy raz jeden raz inny jest lepszy)
4. Eksperymenty z drzewami decyzyjnymi.
  - dokonaj selekcji zmiennych jak w zad. 3
  - zbuduj drzewo decyzyjne, zacytuj / skomentuj dostępne własności drzewa-klasyfikatora
  - przetestuj klasyfikatory uzyskane dla różnych parametrów
  - eksperymenty z przycinaniem drzewa
  - wizualizacja dla dwóch zmiennych decyzyjnych: obszary decyzyjne
5. Grupowanie.
  - potestuj różne metody grupowania, w tym hierarchiczne
  - przedstaw wyniki na wykresach, oznaczając różne grupy różnymi kolorami (w razie potrzeby użyj PCA, lub innej metody redukcji wymiarowości)
  - zbadaj wpływ standaryzacji na grupowanie
  - spróbuj rozstrzygnąć, wszelkimi dostępnymi sposobami, jaka liczba klastrów jest najlepsza
  - użyj metody hierarchicznej do systematyzacji danych w rodzaju “gatunki zwierząt”
6. Reguły asocjacyjne