

Imię i nazwisko:

Grupa (prow. ćwiczenia):

Numer indeksu:

Odpowiedzi powinny być krótkie, najwyżej kilka zdań. W przypadku wyników liczbowych należy w skrócie przedstawić sposób rozwiązania. Zalecane jest zmieszczenie się z rozwiązaniem na niniejszej kartce.

1. Badamy próbkę 10 samochodów, licząc znalezione w nich usterki. Na rysunku obok przedstawiono histogram liczby usterek. Czy na podstawie tego rysunku i powyższych danych można obliczyć średnią liczbę usterek? Medianę? Jeśli tak, oblicz te wartości. Jeśli nie, uzasadnij dlaczego.
2. Podaj przykład próbki (danych liczbowych), dla których kwantyl $q_{0,25}$ (czyli dolny kwartył) równy jest wartości średniej.
3. W regresji liniowej metodą najmniejszych kwadratów wzory na współczynniki regresji dobrane są tak, by minimalizować pewną wielkość. Co to za wielkość?
4. Analizujemy system obsługi sklepu internetowego pod kątem wydajności modułu realizującego autoryzację kart płatniczych. Założenia systemu mówiły o średnim czasie działania tego modułu równym co najwyżej 10 s. Zbadaliśmy rzeczywisty czas działania w 2500 przypadków i chcemy ocenić, czy przy poziomie istotności 2% mieścimy się w założonych ramach. Jak w tym przypadku brzmi hipoteza zerowa? Hipoteza alternatywna? Jaką postać ma zbiór krytyczny? Jakich jeszcze danych potrzebujemy, by zweryfikować hipotezę?
5. Jakie założenie upraszczające przyjęto przy konstrukcji naiwnego klasyfikatora bayesowskiego? (Skąd określenie „naiwny”?)
6. Mamy tablicę decyzyjną złożoną z 500 000 obiektów opisanych 20 atrybutami o wartościach 0 lub 1 (każdy atrybut ma mniej więcej połowę 0 i 1). Chcielibyśmy zanalizować te dane metodą trzech najbliższych sąsiadów (3-nn). Uzasadnij (wykorzystując powyższe dane), dlaczego ta metoda będzie nieskuteczna.
7. Budujemy drzewo decyzyjne, wykorzystując entropię do oceny cięć. Mamy węzeł zawierający 100 obiektów i rozważamy cięcie, które rozdzieli nam 1 obiekt do lewego potomka i pozostałe 99 obiektów do prawego. Jaka będzie entropia lewego potomka? Dlaczego (mimo to) inne cięcia mogą być lepsze?
8. Czym różnią się reguły decyzyjne od reguł asocjacyjnych?
9. Pokaż (na możliwie prostym przykładzie graficznym) jak mogą zmienić się wyniki grupowania algorytmem K-means w zależności od wyboru punktów początkowych.
10. Podaj przykład tablicy decyzyjnej (wartości dyskretne), dla której test niezależności (za pomocą tablicy kontyngencji) wykazuje niezależność od decyzji każdego atrybutu z osobna, natomiast algorytm AQ znalazłby silne i poprawne reguły decyzyjne.

