

Estymacja gęstości i reguła Bayesa (Density estimation and Bayes rule)

Zadania:

1. • Wygeneruj 200- elementową mieszaną rozkładów

$$0.9 * N(5, 1) + 0.1 * N(10, 1).$$

- Wyznacz estymator jądrowy który dość dobrze przybliżałby gęstość teoretyczną f . Narysuj wykresy gęstości mieszanki i estymatora gęstości oraz histogram.
- Oblicz wartość empiryczną błędu średniokwadratowego

$$\frac{1}{K} \sum_{i=1}^K [f(x_i) - \hat{f}_n(x_i)]^2,$$

gdzie $x_i, i = 1, \dots, K$ są równoodległymi punktami podziału odcinka $[2, 12]$.

2. Dane `geyser(MASS)` dotyczą wybuchów gejzerów. Narysuj wykres estymatora gęstości dwuwymiarowej pary zmiennych (**duration, waiting**). Rozstępy wyznacz metodą Shealtera-Jonesa dla każdej zmiennej oddzielnie. Sporządź wykres dla wyznaczonego estymatora gęstości.
3. Dane: w pliku `earthquake.txt`.
 - Wyznacz estymatory gęstości dla zmiennej **body** w obu populacjach. Współczynnik wygładzający ustaw arbitralnie na **bw=0.2**. Sporządź wykresy uzyskanych estymatorów gęstości f_1 i f_2 .
W podobny sposób wykonaj wykresy $\pi_1 f_1$ i $\pi_2 f_2$, uwzględniając prawdopodobieństwa apriori przynależności klasowej π_1, π_2 . Sformułuj postać reguły Bayesowskiej, podaj przybliżoną wartość progu.
 - Wyznacz estymatory gęstości dwuwymiarowych pary zmiennych (**body, density**) w obu populacjach. Sporządź wykresy estymatorów i odpowiadające im wykresy konturowe, zastosuj funkcje **persp** i **contour**.
4. Na wybranym zbiorze danych porównaj działanie Naiwnego Klasyfikatora Bayesa, metody LDA i QDA.