

Density estimation

Tasks (Lab 6):

(all tasks are scored)

1.
 - Generate sample of size 200 from the mixture of two distributions:

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

- Use **kernel density estimator** to approximate the theoretical density f corresponding to the above distribution. Draw density function corresponding to theoretical distribution and estimated function \hat{f}_n .
- Compute mean squared error

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

where $x_i, i = 1, \dots, K$ are equally distributed points from interval $[2, 12]$.

- Analyse how the error depends on sample size n .
 - Analyse the influence of different kernel functions as well as smoothing parameters.
2. Select any dataset corresponding to binary classification problem with quantitative variables. Compare the accuracy of Naive Bayes method (with kernel density estimator), Naive Bayes Method (with Gaussian approximation), Naive Bayes method (with discretization of quantitative features) and LDA.