Density estimation, naive Bayes Method

Tasks:

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, $i = 1, \ldots, K$ 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. Consider simple dataset *earthquake.txt* and simple classification rule based on one variable (body).

   • Estimate density functions of variable **body** in both classes (call the f_1 and f_2). Draw the curves of f_1 i f_2.

   • Draw the curves of $\pi_1 f1$ i $\pi_2 f2$, taking into account apriori probabilities π_1, π_2. Find the threshold corresponding to Bayes classification rule.

3. Select any dataset corresponding to binary classification problem wih 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.