Density estimation

Tasks (Lab 3):

1.
   - Generate sample of size 200 from the mixture of two distributions:
   $$0.9 \cdot N(5,1) + 0.1 \cdot N(10,1).$$
   - 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 (MSE)
   $$\frac{1}{K}\sum_{i=1}^{K}[f(x_i) - \hat{f}_n(x_i)]^2,$$
   where x_i, $i = 1, \ldots, K$ are equally distributed points from interval $[2, 12]$.
   - Analyse how the error depends on sample size n. Generate a plot showing how the mean squared error changes with n.
   - Analyse the influence of different kernel functions as well as smoothing parameters. Generate curves for different kernels and different values of smoothing parameters. Try at least 3 kernels and 3 values of smoothing parameters. In addition, use one of the smooting parameter estimation methods.
2. Generate sample $X_1, \ldots, X_n$ of size $n = 200$ from Gauusian mixture model described above. Compare two methods (compute MSE for both of them):
   (a) **Method 1:** Kernel density estimator using sample $X_1, \ldots, X_n$
   (b) **Method 2:** Using sample $X_1, \ldots, X_n$, generate artificial sample $X'_1, \ldots, X'_k$ (you can choose k much larger than n) corresponding to the kernel density as follows:
      - Generate i from uniform distribution on $\{1, \ldots, n\}$
      - Generate ϵ from $N(0, 1)$
      - Set $X'_i = X_i + \epsilon \cdot h$

      Compute kernel density estimator using modified sample $X'_1, \ldots, X'_k$.
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 of the density)
   - Naive Bayes method (with discretization of quantitative features)
   - LDA