Regression I (scored tasks: 4 points)

Tasks (Lab 11):

1. Generate artificial data using the following equation:

$$y = g(x) + \varepsilon,$$

where: $g(x) = 4.26(e^{-x} - 4e^{-2x} + 3e^{-3x})$, and ε follows normal distribution with parameters $\mu = 0$, $\sigma = 0.1$. The above function is a benchmark function frequently used to test advanced regression methods. Assume that x_i are generated from uniform distribution on $[0, 4]$.

2. Implement from scratch Nadaraya-Watson kernel regression estimator.

3. Compare the performance of the two algorithms:

   - Nadaraya-Watson kernel regression,
   - Smoothing Splines regression.

4. Visualize the results. Prepare scatter plot and draw the curves corresponding to the true regression function $g(x)$ and to the fitted models.

5. Compute $MSE = n_{test}^{-1} \sum_{i=1}^{n_{test}} [g(x_i) - \hat{g}(x_i)]^2$ on the test set for the two methods considered. As a test set, use a sample independent from training sample.

6. Explore how the MSE depends on the training sample size n. Generate a plot.