Laboratory 6 – Exercises

1. Generate the following dataset:

$$y_i = -4\cos x_i + 3x_i + \epsilon,$$

where ϵ obeys a normal distribution with mean 0 and standard deviation 0.08. The number of training/testing examples is 100. Plot the prediction for this regression problem using the Nadaraya-Watson kernel regression non-parametric attention model, and also plot the attention heatmap.

2. Train a Transformer model for machine translation. The dataset is the English-French dataset from the laboratory. Modify the `PositionWiseFFN` class by using 3 hidden layers of 32 units. The dropout probability will be set to 70%. Train the model for 400 epochs and plot the training loss.