Assignment #6

Course: *Machine learning*
Due date: *December 15th, 2023*

Assignment

In this assignment, you will learn about kernel methods.

Download the Breast Cancer Wisconsin dataset `https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data` and perform all the necessary data pre-processing.

Use the Scikit-learn implementation of SVM (the SVC() function) to predict if the tumor is malignant or not. Try different kernels.

Try different:

- kernels
- regularization parameters
- different σ values when using the RBF kernel (in Scikit-learn this is the γ parameter).

  Find the best values for the parameters and do a grid search. Tip: take a look at the GridSearchCV() function.

Test your models.

Be careful not to use the same data for any stage of training and testing your models.

Implement kernel regression to model the data from Učilnica.

Train and test your kernel regression model.

How do I implement a kernel regression?

- Implement at least three different kernel functions.
- Compute the kernel matrix.
- Compute $\alpha = K^{-1}y$ to get the model coefficients in the multidimensional space.
- Compute predictions for the test point z: $y(z) = k_*\alpha$. Where k_* is the kernel (vector) of the test points with the training points.

$$k_*(z) = \begin{bmatrix} k(z, x_0) \\ \vdots \\ k(z, x_n) \end{bmatrix}$$

Plot the data and your predictions.