Instructions

- This assignment consists of 4 exercises:
    - Use ANFIS to solve the regression problem (Dataset 1);
    - Use Neural Networks to solve the regression problem (Dataset 1);
    - Use ANFIS to solve the classification problem (Dataset 2);
    - Use Neural Networks to solve the classification problem (Dataset 2).
- Use Python, MATLAB or other language to solve the assignment.
- Deadline for submission of the report is **October 3rd, 2025 at 23:59**.
- Deliver all data files in GitHub, where the best models that you have developed should be included. Deliver also a script file in the language that you have chosen, which performs the required steps to obtain the developed model. Further, deliver a file ``******_A2_report.pdf`` (****** should be substituted by your student number) in which you explain how the model parameters were selected and present the main results, elaborating a suitable discussion and comparison between all the used methods, including the methodology used in assignment 1. This report can be a notebook or a live script as long as it includes the above explanations and is converted to pdf. Make sure that your name and student number are indicated clearly.
- The report **must** be submitted in the Fenix system and it is **mandatory** to include the link to the work on GitHub. **Remember to make your repository public!!!**

Dataset 1: Diabetes Dataset (Regression)

Task

Predict a quantitative measure of disease progression one year after baseline using 10 baseline variables:

- age – age in years
- sex – gender (binary)
- bmi – body mass index
- bp – average blood pressure
- s1 – total serum cholesterol

- s2 – low-density lipoproteins
- s3 – high-density lipoproteins
- s4 – total cholesterol / HDL
- s5 – serum triglycerides
- s6 – blood sugar level

Target

y - A quantitative measure of diabetes progression one year after baseline.

Dataset size

- 442 samples
- 10 features
- 1 target

Link for the dataset:

https://www.kaggle.com/datasets/antoniofurioso/diabetes-dataset-sklearn

Alternatively, (if you are using python) you can import the dataset directly from scikit-learn:

```python
from sklearn import datasets

# Load dataset
diabetes = datasets.load_diabetes( as_frame=True)
X = diabetes.data
y = diabetes.target
```

Dataset 2: Pima Indians Diabetes Dataset (Classification)

Task

Predict whether a patient shows signs of diabetes based on 8 diagnostic measurements:

- Pregnancies – number of times pregnant
- Glucose – plasma glucose concentration (2 hours in oral glucose tolerance test)
- Blood Pressure – diastolic blood pressure (mm Hg)

- Skin Thickness – triceps skinfold thickness (mm)
- Insulin – 2-hour serum insulin (µU/ml)
- BMI – body mass index (kg/m^2)
- Diabetes Pedigree Function – a function capturing family history of diabetes
- Age – age in years

Target:

y - Binary outcome: 1 (diabetes) or 0 (no diabetes).

Dataset size

- 768 samples
- 8 features
- 1 target

Link for the dataset:

https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

Alternatively, (if you are using python) you can import the dataset directly from scikit-learn:

```python
from sklearn.datasets import fetch_openml

# Load dataset
diabetes = fetch_openml("diabetes", version=1, as_frame=True)
X = diabetes.data
y = diabetes.target
```