Machine Learning
Laboratory no. 5: Breast Cancer Diagnostic

Aim

- Download and analyse the dataset
- Train one or more machine learning algorithms in order to make predictions and classify new samples
- Evaluate the model performance

1 Instructions

This laboratory will be evaluated. At the end of the laboratory, you will upload to Blackboard your individual work consisting of (i) your structured code (`.py` or `.ipynb`) with comments, and (ii) your report (`.pdf`) in English. The report will incorporate an introduction and a conclusion. It will describe your work, analyse the results, present the difficulties that you could encounter and description how you solved them.

2 Database

The database contains several attributes related to breast cancer. The aim is to predict for a patient whether his/her attributes are benign or malignant.

To use it, type the following lines:

```
from sklearn.datasets import load_breast_cancer
samples = load_breast_cancer()
```

We consider that attributes are saved as the variable x and that the corresponding classes are saved as the variable y.

```
import pandas as pd
x = pd.DataFrame(samples.data)
y = pd.DataFrame(samples.target)
```

3 Advice

- Analyse the data. For example, how many samples are there?
- Prepare the data before training a model
- Choose one or several models that are suitable for the studied case

- Evaluate the model performance
- Comment the code
- Note down your observations, analyses and comments in the report
- Plot the data to visualise them
- Cite the resources that you use

4 Indicative Grading

	Number of points
Compliance with the rules	1
Code	5
Introduction	1
Description and justification of methods	4
Raw results	4
Result analysis	4
Conclusion	1