

# 1 Classification data

In this task, we are going to study the well-known Iris data set of the UCI Machine Learning repository.

1. Describe the data set in terms of inputs, outputs, and training data.
2. Identify the machine learning task connected to it.
3. Load the data set into the provided empty Jupyter notebook such that the categorical outputs are delivered as labels  $\{1, 2, \dots\}$ .

## 2 Regression data and representation generation

In this task, we consider much more involved data, namely molecular data. You will be provided a demonstration of how to characterize the data and how to produce a proper representation for its use in machine learning.