

CO395 - Dataset description

Prepared by Antoine Cully

The best way to know about machine learning is to try and test it by yourself. Two simple datasets are provided here that enables you to test and apply the different machine learning algorithms seen in class.

1 IRIS dataset

The IRIS dataset is a simple dataset (https://en.wikipedia.org/wiki/Iris_flower_data_set, provided in the file `iris.dat`). It can be used for classification and clustering tasks. The dataset contains a matrix in which each line is a different sample. There are 150 records under five attributes - petal length, petal width, sepal length, sepal width and species. The species are encoded with a one-hot encoding and can have 3 different values (thus encoded as 3 dimensions).

2 Robot dataset

The Robot dataset is used to predict the position of an industrial robot given the motor commands (this is what we usually call a forward model). Here, the robot is an ABB IRB 120, which is a very popular industrial robot (see figure 1), used in many assembly lines.

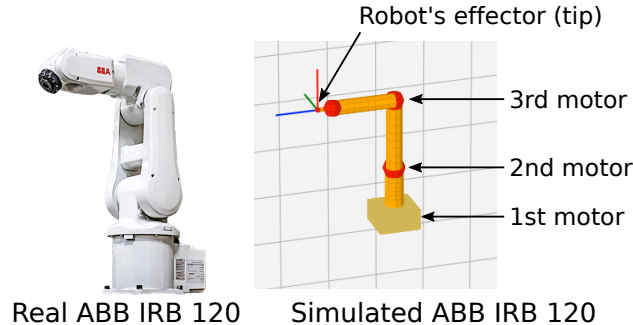


Figure 1: ABB IRB 120

While this robot has 6 degrees of freedom (i.e., 6 inputs), the dataset only considers the first three degrees of freedom, which control the final position of the robot's tip (end-effector). The dataset can be used to predict the Cartesian position (x/y/z) of the robot's tip given the angular position of the first three motors ($\theta_1, \theta_2, \theta_3$). Such a model is usually called in robotics the “forward model” of the robot.

The dataset *robot.dat* contains a matrix in which each line is a different sample. The first three columns correspond to the inputs (angular positions of the motors), and the last 3 ones correspond to the Cartesian position of the robot's tip (X/Y/Z position in mm).