Artificial Intelligence 2020/2021

Exercise Sheet 5b: Supervised Learning

5.3 Iris flower extended data set –Classification using different Algorithms

Continuing with the Iris dataset, suppose that we have Iris already identified in the 3 classes but now we have also the Iris packed in different types of packages: “Simple – 0”, “Gift – 1” and “Luxury – 3”. We also have a new variable “price” with three possibilities: “Low”, “Medium”, “High”.

Iris setosa | Iris versicolor | Iris virginica

We have now a different classification problem in which we want to predict the “price” classification based on the remaining characteristics: sepal_length_cm, sepal_width_cm, petal_length_cm, petal_width_cm, iris_type, and package.

a) Create a new notebook and start by importing the needed libraries.

b) Read the data from the CSV file and check the data using the head(), describe(), and other Pandas commands.

c) Using only the attribute sepal_length_cm, sepal_width_cm, petal_length_cm, petal_width_cm, fit a simple decision tree model to the data.

d) Analyse the accuracy, precision, recall and f-measure achieved.

e) Create and analyse a confusion matrix of the results.

f) Using only the attribute sepal_length_cm, sepal_width_cm, petal_length_cm, petal_width_cm, fit a simple nearest neighbor model to the data.

g) Analyse the accuracy, precision, recall and f-measure achieved.

h) Fit distinct models such as decision trees, SVM and neural networks to the data and try different configuration parameters.

i) Analyse the accuracy, precision, recall and f-measure achieved for distinct models.