Information Theory and Data Science

-

Assignment no. 3

-

Mauro De Sanctis

General Recommendations

The project should be carried out using Python. Good programming discipline should be followed when writing the Python code. This means that the variable names should be logical, the code must be commented and it should be written in such a way that it is easy to follow and understand.

Project

- Compute the probability mass function of the features of the Iris dataset after having discretized the samples as integers.

- Compute the entropy of the features of the Iris dataset (discrete entropy).

- Compute the mutual information between any pair of features of the Iris dataset (discrete mutual information).

Note: Start from the example in pmf_estimation_multivariate.py and test_pdf_multivariate_iris.py