Information Theory and Data Mining

-

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 difference between the entropy of a discrete random variable given its p.m.f. vector and the entropy of an estimated p.m.f. from a set of samples generated through the same p.m.f. vector. Choose freely one p.m.f. vector.

- Write a function called "differential_entropy" which computes the differential entropy of a generic continuous random variables given its p.d.f.

- Compute the difference between the differential entropy of a Gaussian continuous random variable given its p.d.f. vector and the differential entropy of an estimated p.d.f. from a set of samples generated through the same p.d.f. vector. Choose freely the mean and variance.

Note: consider the impact of bandwidth, kernel function and number of samples on the difference between such entropies. Start from the example in pmf_estimation.py and test_pmf_pdf.py