Information Theory
Assignment 1: Entropy

Deadline: 15.09.19 23:59
Output: ZIP-file with the code and PDF-report has to be uploaded to Moodle, input files are NOT needed. Name of the archive has to be *NameSurname.zip* (For example, *IvanIvanov.zip*). Name of the source code has to be *NameSurname.py* (For example, *IvanIvanov.py*) Name of the report has to be *NameSurname.pdf* (For example, *IvanIvanov.pdf*). No other symbols allowed
Programming language: Python 3.7
Requests:

- The program must work, the code should be readable, well-structured and should contain English comments
- NO extension of a deadline. Works sent after the deadline will NOT be evaluated
- Assignment is strictly individual
- We will be using MOSS (Measure of Software Similarity) as a test for plagiarism. Be reminded that a score of 0 will be assigned to any submissions suspected of plagiarism pending a full investigation as per IU policies.

Evaluation criteria:
0 (0%) - no submission or late submission
1 (20%) - required functionality is not achieved
2 (40%) - required functionality is lower than 50%
3 (60%) - required functionality is 50-80% and/or shortcomings in report
4 (80%) - required functionality is 80-100% and/or shortcomings in report
5 (100%) - well-structured readable correct code with English comments and correct report
Task:
Write a program that will read each file and calculate the average entropy for different types of files. We consider file as a sample of byte distribution. Each byte can take a range of values (guess which) with some probabilities. The task is to estimate those probabilities and to compute the file entropy (relative to the file length) based on that data. Consider the following file types:

1. Picture (jpeg, png)
2. Executable file (exe or Linux/macOS executable)
3. Text (doc, pdf). Use only textual data inside

Compare the average entropy for all types of files in report. Does it differ for the files of the same type? Does it differ for the files of different type? Compare obtained results with maximal entropy for that range of values. Explain your answers