

Assignment 0

August 2020

The objective of this lab assignment is to familiarize you with the `numpy` and `matplotlib` libraries in `python`.

We need to give clear instructions on how to submit and what/format to submit. That was a constant complaint last year.

1 Problem on Probability

1.1 Theoretical Problems

- (a) Let $X \sim F_X$ be a random variable with given CDF F_X . Let there be another random variable $Y \sim \text{Unif}[0, 1]$. Express X in terms of Y and prove your result.
- (b) Let X be an exponential distribution with parameter λ . Using the above result, obtain X in terms of Y .

1.2 Programming

Let X be from the exponential distribution with $\lambda = 1.0$. Complete the function `modify_uniform` in the file `probability.py` to generate samples of this variable using only the `numpy.random.uniform` function.

2 Vectorization

In this problem you will implement a function to compute the pair-wise L_2 similarity between each pair of points in a set.

Let $X \in \mathbb{R}^{n \times d}$ where n is the number of points in the the set and d is the number of dimensions of the basis if the points, the L_2 similarity between two points x and y is defined as

$$d(x, y) = \sum_{i=1}^d (x_i - y_i)^2$$

- (a) Obtain a vector expression for $d(x, y)$ when $x, y \in \mathbb{R}^d$

- (b) Complete the function `pairwise_similarity_looped` in the file `similarity.py` to obtain this matrix K using for loops.

The above $\mathcal{O}(n^2d)$ solution doesn't scale well. However, `numpy` provides a powerful mechanism called vectorization which can speed up this process drastically.

- (c) Complete the function `pairwise_similarity_vec` in the file `similarity.py` to obtain K in a vectorized manner. Refer to the comments in the function for an approach to this problem.
- (d) Run the file `similarity.py` for multiple values of d and n using the command

```
$ python3 similarity.py --dim <d> --num <n>
```

How do you expect time take for the vectorized and looped functions to grow with dimension and number of samples? Plot 4 graphs showing the time vs dimension and time vs number of samples for the two functions and include these in your solutions file.

3 Probability and simulation

You are given a special coin for which probability of getting a Head is 0.25 and probability of getting a Tail is 0.75. You are told to keep flipping the coin till you get two consecutive heads. What is the expected number of flips that you have to make?

- (a) Compute the expected value analytically.
- (b) Write a `numpy` program in the file `simulation.py` to simulate this experiment for $n = 10, 100, 1000, 10000$ steps and plot a graph of the observed expected value vs number of steps. Report your observations.

Submission Instructions