

Yaffa Tziporah Atkins

## Homework 1

### Machine Learning

#### 1. Solutions

Compute distance loop version: Used Pythagorean Theorem. I compared two rows to find their differences through summing the squares of the differences in each column.

Compute distance vectorization version: used basic NumPy math to perform the function for me, as opposed to manually doing it myself.

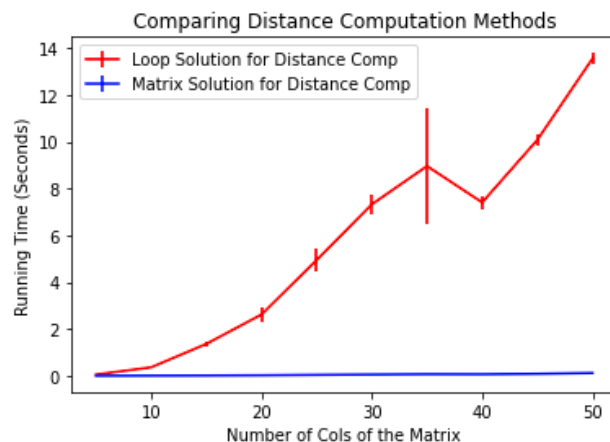
Compute correlation loop version: I used the correlation formula. I subtracted the average of the column from each element, so the average is zero. I looped through each pair of columns and took the sum of their products and divided it by the product of the square roots of the sum of the products.

Compute correlation vectorization version: I used functions to find the correlation for me.

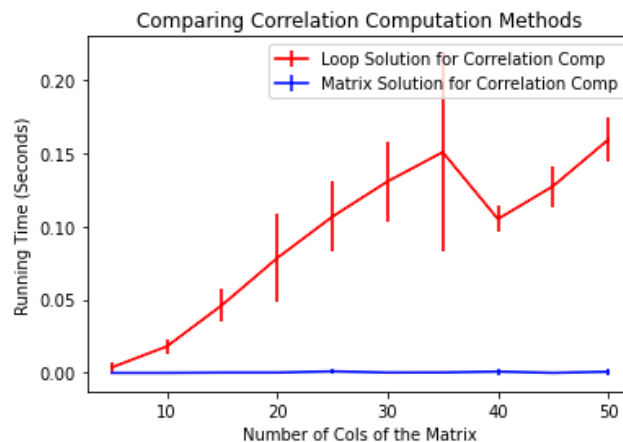
The dimensions I used for the random matrices are (50,5), (100,10), (150,15), (200,20), (250,25), (300,30), (350,35), (400,40), (450,45) and (500,50)

#### 2. Experiments:

##### Problem 1:



##### Problem 2:



Problem 3:

	Naive	Smart
Iris	0.169546	0.001995
Breast Cancer	10.911746	0.109108
Digits	216.547214	2.014653
	Naive	Smart
Iris	0.000997	0.000000
Breast Cancer	0.067822	0.000998
Digits	0.545606	0.002992

Size is the control variable for this experiment. I observed that the larger the sample set, the exponentially more time the methods take. But when you compute the function without loops, using vectorization, the solution is calculated much faster.