**Fundamentals of Data Science**

**Coding Project**

Veera Raghunatha Reddy Naguru – 22028322

I have been given a dataset named ‘data2.csv’, which contains 400 weights of new born babies born in regions of Europe as part of this project. It is asked to read the data as an array, create a distribution of that array, find average weight W of that distribution and X such that X is fraction of weights between W and 1.2\*W.

I used numpy and matplotlib (pyplot) in python to solve this project. To read data as an array I used np.genfromtxt which is from numpy module. I have used np.histogram to create distribution. The distribution is spread around weights ranging from 1.86Kg to 4.83Kg with 13 equal widths.

The resulted distribution is used to create distribution centres by taking averages of the distribution edges. The distribution frequencies are divided by total number of weights to normalise the distribution, by this we get probability distribution function of weights in the distribution.

The resulted array of distribution centres and probability distribution function is given below:

Distribution Centres –

[1.98104962, 2.20916885, 2.43728808, 2.66540731, 2.89352654, 3.12164577, 3.349765, 3.57788423, 3.80600346, 4.03412269, 4.26224192, 4.49036115, 4.71848038]

Probability Distribution Function (PDF) –

[0.02, 0.02, 0.05, 0.0975, 0.1475, 0.135, 0.1675, 0.1425, 0.13, 0.0625, 0.02, 0.0025, 0.005]

Using the above distribution centres and PDF we can calculate average weight W by summing the product of distribution centres and pdf. Appropriate formulae for this can be notated as below:

**W=∑ (Distribution Centres \* Probability Distribution Function)**

By doing so I got Average Weight **W = 3.25Kg.**

Similarly for X, we can sum the probability distribution function where distribution centres lie in between W and 1.2\*W. This can be notated as below:

**X=∑ PDF (W < Distribution Centres < 1.2\*W)**

By doing so I got fraction of weights between W and 1.2\*W – **X = 0.44.**

After getting all the outputs, I used matplotlib.pyplot bar,plot and text to display the histogram and other outputs in a single plot. Appropriate labels, title and legend are given.