جامعة الطائف
TAIF UNIVERSITY

PATTERN CLUSTERING WITH KOHONEN SELF-ORGANIZING MAPS (SOM)

Neural Networks

Presented to: Dr Huda Hakami

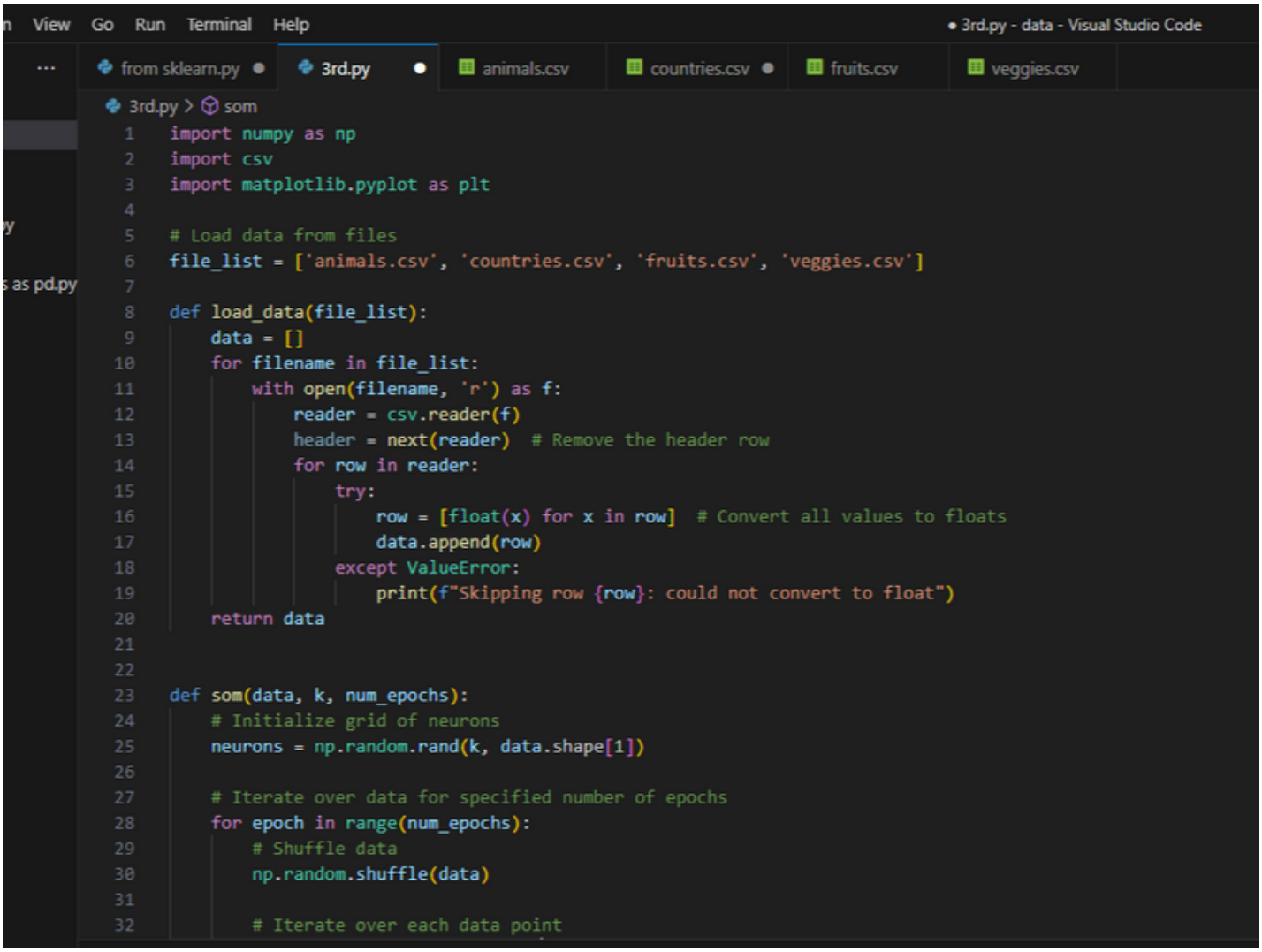
WORD CLUSTERING USING SOMS

student name	number
Ebtsam twim Alyzidi	43903763
Manal Abdullah Alghamdi	43907535
Nouf Ahmed Altowairqi	43900958

Questions/Tasks :

(1) Write a program to load the data instances to memory from the provided data file.

We have converted the data file into a csv file, using the csv library, and loading the data, also converting the data into floats so that it can be read.



(2) Implement the SOM algorithm to cluster the training instances (i.e., words). Train SOM with k neurons (i.e., clusters).

The SOM algorithm involves creating a grid of neurons, where each neuron represents a cluster. During training, each data instance is presented to the SOM, and the neuron closest to the input data is selected as the winner. The weights of the winning neuron and its neighboring neurons are updated to move them closer to the input data. This process is repeated for all input data instances until the weights of the neurons converge.

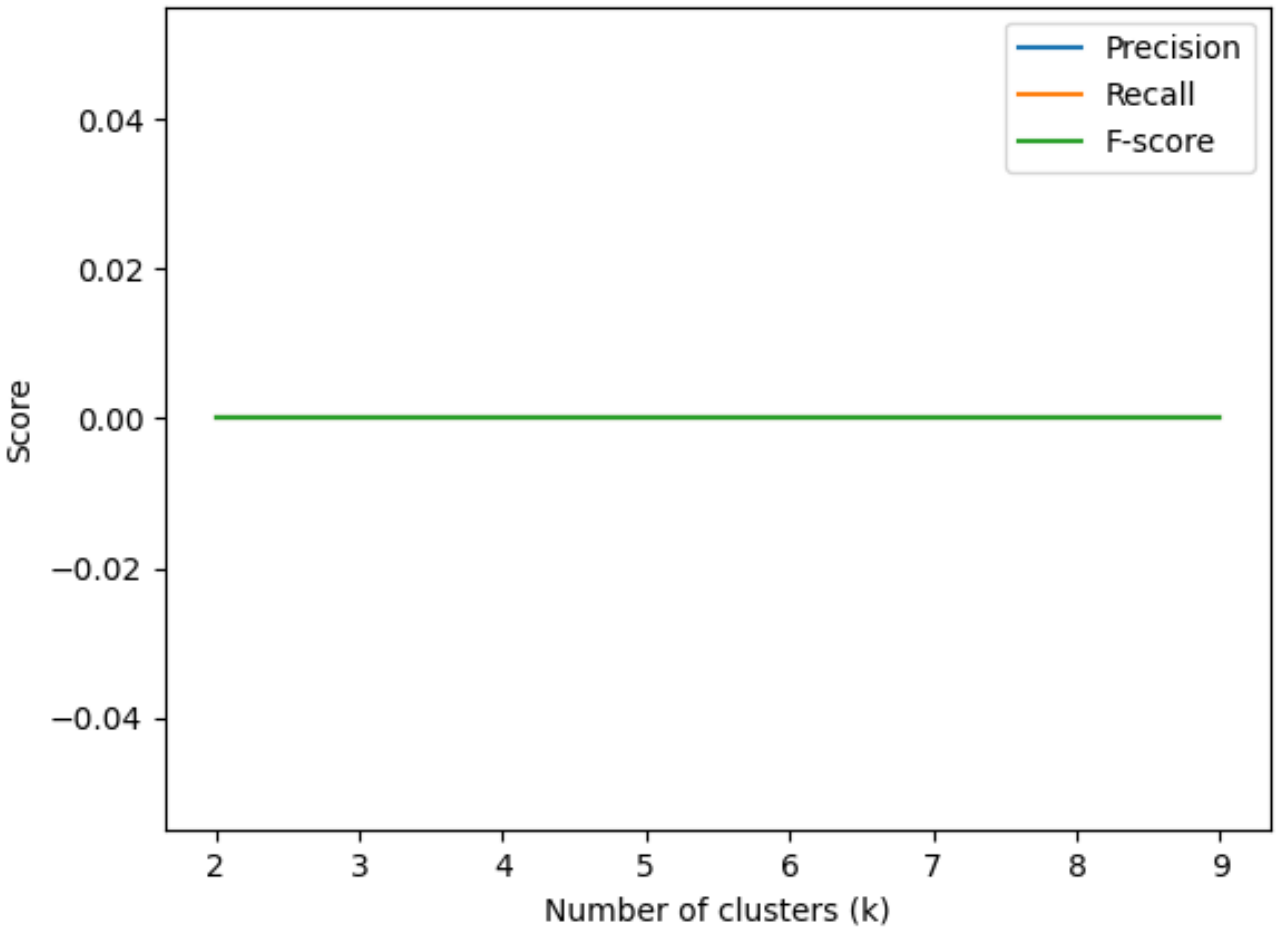
(3) Vary the value of k (neurons) from 2 to 10 in the SOM and compute the precision, recall, and F-score for each set of clusters. Plot a figure that shows k in the horizontal axis and precision, recall and F-score in the vertical axis in the same plot

To do this, you'll need to run the SOM algorithm multiple times with different values of k, and then evaluate the resulting clusters using precision, recall, and F-score. These metrics can be calculated by comparing the clusters produced by the SOM with the ground truth labels for the data instances. You can use a clustering evaluation metric such as the Adjusted Rand Index (ARI) or Normalized Mutual Information (NMI) to compute these metrics.

This can be done using a plotting library such as Matplotlib in Python. You can plot a separate line for each metric, with k on the x-axis and the metric value on the y-axis.

But we had a problem and unfortunately we will not be able to solve it. We have tried a lot, but the values have not changed, all of them are zero

Here the representation of the values has never changed, even after doing a lot of modifications to the code



References:

https://machinelearningmastery.com/implement-backpropagation-algorithm-scratch-python/

https://www.google.com.sa/books/edition/Machine_Learning_and_Big_Data_Analytics/qfsOEAAAQBAJ?hl=ar&gbpv=0

https://towardsdatascience.com/text-clustering-using-k-means-ec19768aae48