CS 534 Homework Assignment 4 (100 Points)
Due: 11:59 p.m. on 07/10/2022

Homework Objective: The goal of this assignment is to help you master the materials of Lecture 6 and 7.

Homework Deliverables: Submit your answers and solutions in a *zip* file that includes all the write-ups in a *pdf* file and your *python codes* to Canvas.
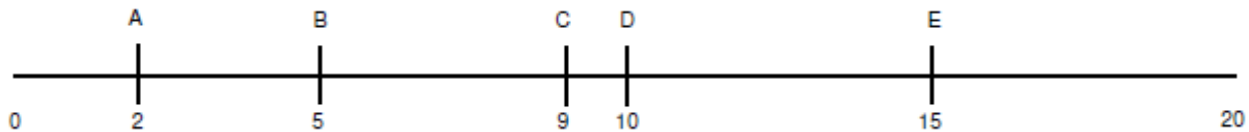
Notes:
1. Each answer needs to be explained.
2. The implementation should be in Python if You need.

Project Questions:

Question 1: *K*-means Clustering

Five Customers' Rating on a New Car on a 20-point Scale.

(a) Assume $K = 2$ and the two initial centroids are 3 and 4.
1. Use the *K*-means algorithm and show all the computational steps with the numerical answers and solution to determine the two-cluster solutions. **No point** is given if only the answer is provided.

2. Show all the computational steps with the numerical answers and solution to calculate Silhouette Coefficient Index, Davies–Bouldin Index, and Calinski-Harabasz Index. **No point** is given if only the answer is provided.

(b) Assume $K = 2$ and the two initial centroids are 11 and 12.
1. Use the *K*-means algorithm and show all the computational steps with the numerical answers and solution to determine the two-cluster solutions. **No point** is given if only the answer is provided.

2. Show all the computational steps with the numerical answers and solution to calculate Silhouette Coefficient Index, Davies–Bouldin Index, and Calinski-Harabasz Index. **No point** is given if only the answer is provided.

(c) Use the results from (a) and (b) to determine which two-cluster solution should be chosen. Please describe and explain your answer in detail.

Question 2: Association Analysis

Consider the following customers' transactions in a supermarket.

Transaction	Milk (M)	Eggs (E)	Bread (B)
Customer 1	0	1	1
Customer 2	1	0	1
Customer 3	1	1	0
Customer 4	0	1	1
Customer 5	1	1	1
Customer 6	0	1	1
Customer 7	1	1	1

(a) Generate all the possible itemsets and fill in the below table.

Itemset	Size of the Itemset	Frequency of the Itemset	Support

(b) Suppose the support threshold is 30%. Generate all the possible association rules from the frequent itemsets and fill in the below table for each rule. You only need to consider the frequent itemset if the size of it is at least 2.
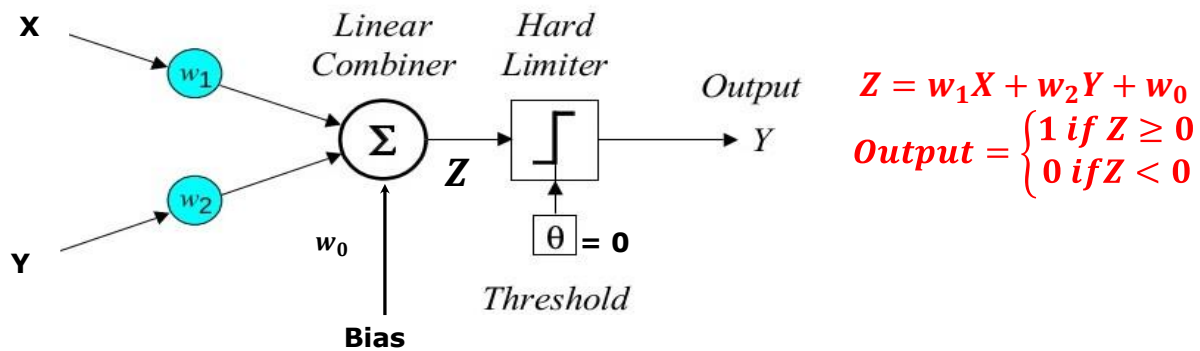
Association Rules		Confidence
Antecedent	Consequent	

(c) Write down all the association rules if the confidence threshold is 70%.

Question 3: Deep Learning

Train a perceptron to classify the following instances correctly.

Inputs

X

w_1

Linear Combiner

Hard Limiter

Σ

Z

w_2

Y

w_0

Bias

$\theta = 0$

Threshold

Output

Y

$Z = w_1 X + w_2 Y + w_0$

$$Output = \begin{cases} 1 \ if \ Z \geq 0 \\ 0 \ if Z < 0 \end{cases}$$

X	Y	Output
0	0	0
0	1	1
1	0	1
1	1	1

Assume that initial $w_1 = 0.75$; $w_2 = 0.45$; $w_0 = -0.75$. **Z = 0.75X + 0.45Y − 0.75**. Learning Rate = 0.1. Show all the steps to demonstrate the learning process and the final perceptron equation. **No points** are given if only the answer is provided.

Question 4: Convolutional Neural Network

(a) Consider a 1-dimensional time-series with values 2, 1, 3, 4, 7. Perform a convolution with a 1-dimensional filter 1, 0, 1 and zero padding. Show all the computational steps to get the full credit.

(b) Compute the convolution of the below input with the horizontal edge detection filter. Use a stride of 1 without padding. Show all the computational steps to get the full credit.

INPUT

1	1	1
0	0	0
-1	-1	-1

HORIZONTAL EDGE DETECTING FILTER
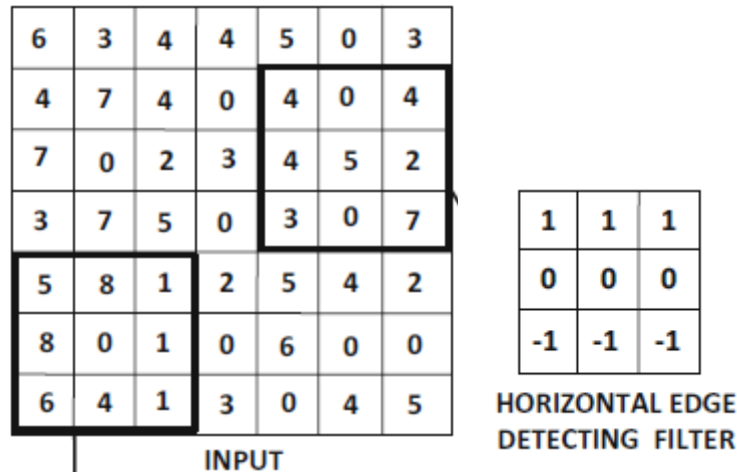
(c) Perform a 4 × 4 pooling at stride 1 of the above input. Show all the computational steps to get the full credit.

Question 5. AlexNet Architecture

(a) Download an implementation of the AlexNet architecture from a neural network library of your choice (e.g., https://github.com/paniabhisek/AlexNet).

(b) Train the network on 100, 150, 200, 250, and 300 *color plant images*, respectively, of your choice from https://github.com/spMohanty/PlantVillage-Dataset/tree/master/raw/color, and plot the 5 errors with those data sizes in a line chart. Please submit your code and output the results.

Grading Criteria: Your answers must be complete and clear.