BINUS University

Academic Career: ***Undergraduate / ~~Master~~ / ~~Doctoral~~ *)***			**Class Program:** ***~~International~~/Regular/~~Smart Program~~/~~Global Class~~*)***		
☐ **Mid Exam** ☑ **Final Exam** ☐ **Short Term Exam** ☐ **Others Exam : _____**			**Term : ~~Odd~~/Even/~~Short~~ *)**		
☑ **Kemanggisan** ☑ **Alam Sutera** ☑ **Bekasi** ☐ **Senayan** ☐ **Bandung** ☐ **Malang**			**Academic Year :** **2020 / 2021**		
Faculty / Dept.	:	School of Computer Science	Deadline	Day / Date :	Friday / Jul 23rd, 2021
				Time :	17:00
Code - Course	:	COMP6577 - Machine Learning	Class	:	All Classes
Lecturer	:	Team	Exam Type	:	Online
*) *Strikethrough the unnecessary items*					
The penalty for CHEATING is DROP OUT!!!					

Learning Outcomes:

LO 1 : Explain the fundamental of machine learning concept
LO 2 : Interpret the distribution of dataset using regression method
LO 3 : Experiment classification and clustering algorithm from given dataset

I. ESSAY (75%)

1. **(LO 1, LO 3, 20 points)** Observe thoroughly the table below showing 4 datasets with 10 dimensions. If your object query having vector (1,1,1,1,1,1,1,1,1,1) and the similarity is calculated based on Euclidean distance metric and k-NN (k-Nearest Neighbour) algorithm, find the most similar object with the object query. To obtain a reasonable computational accuracy, make sure the dataset has been normalized and you get the best possible "k" (remember the elbow curve).

ID	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
1	1.1	100	1.2	1.6	1.6	1.1	1.2	1.2	1	1
2	1.4	1.4	1.4	1.5	100	1.4	1.2	1.2	1	1
3	1	1	1	1	1	1	2	100	2	2
4	20	20	20	20	20	20	20	20	20	20

2. **(LO 1, LO 3, 25 points)** Given datasets (UAS_ML2021_DATASET01.csv), determine weight *W's* and bias *b* of the corresponding SVM linear discriminant function $f(X) = W^T . X + b$ using **hand-calculation**. You may need to modify the label/target values from 1 and 0 become 1 and -1. To confirm whether your calculation result is correct, you may run the corresponding SVM library provided by python (write the code in python notebook please). Make sure your SVM parameters fulfill the following the constraints $K(X_i, X)$ linear Kernel:

Verified by,

[Ajeng Wulandari] (D6422) and sent to Program on June 23, 2021

Constraint 1 → $\sum_i^m \alpha_i y_i = 0$

Constraint 2 → $\sum_i^m \alpha_i y_i K(x_i, x) + b = +1$

Constraint 3 → $\sum_i^m \alpha_i y_i K(x_i, x) + b = -1$

3. **(LO 1, LO 3, 30 points)** Given 2D datasets (UAS_ML2021_DATASET02.csv) and assuming you do understand the concept of unsupervised learning, please carry out the below tasks as thorough as possible:

A. Find the best possible "k" used for GMM clustering algorithm using **silhouette scores.**
B. After initializing **means (μ_i), covariance matrix (Σ_i), and phi (π_i),** demonstrate how to calculate responsibilities/weights using E-step of the GMM algorithm and the associated log-likelihood.
C. Using M-step of the GMM algorithm, now update the values of means, covariance and phi of each cluster.
D. Go back to B, calculate new responsibilities/weights and new log-likelihood. Check if there is any convergence of the log-likelihood values.
E. To confirm your results you may need to call the corresponding GMM python library from scikitlearn. Visualize the obtained clusters.

Verified by,

[Ajeng Wulandari] (D6422) and sent to Program on June 23, 2021

II. CASES (25%)

4. **(LO 1, LO 2, LO 3, 25 points)** Based on three (3) variables input namely gmat, gpa, and work_experience, you are expected to implement a logistic regression based model that can be employed to assist an admission process in a company where you are working for. The model may be trained and tested using the following datasets respectively:

gmat	gpa	work_experience	admitted
780	4	3	1
750	3.9	4	1
690	3.3	3	0
710	3.7	5	1
680	3.9	4	0
730	3.7	6	1
690	2.3	1	0
720	3.3	4	1
740	3.3	5	1
690	1.7	1	0
610	2.7	3	0
690	3.7	5	1
710	3.7	6	1
680	3.3	4	0
770	3.3	3	1
610	3	1	0
580	2.7	4	0
650	3.7	6	1
540	2.7	2	0
590	2.3	3	0
620	3.3	2	1
600	2	1	0
550	2.3	4	0
550	2.7	1	0
570	3	2	0
670	3.3	6	1
660	3.7	4	1
580	2.3	2	0
650	3.7	6	1
660	3.3	5	1
640	3	1	0
620	2.7	2	0
660	4	4	1
660	3.3	6	1
680	3.3	5	1
650	2.3	1	0
670	2.7	2	0
580	3.3	1	0
590	1.7	4	0
690	3.7	5	1

Table 1. Training dataset

Verified by,

[Ajeng Wulandari] (D6422) and sent to Program on June 23, 2021

gmat	gpa	work_experience	admitted
590	2	3	….
740	3.7	4	….
680	3.3	6	….
610	2.3	1	….
710	3	5	….

Table 2. Testing dataset

Prior to testing the model, make sure the training dataset above is split for training and validation. Please check the performance of the model using CF (confusion matrix) and in accordance to the obtained CF, give your brief evaluation of the model. You are supposed to write the code from scratch in python notebook.

-- Good Luck --

Verified by,

[Ajeng Wulandari] (D6422) and sent to Program on June 23, 2021