CSCI3230 / ESTR3108: Fundamentals of Artificial Intelligence 2023-2024 Term 1
The Chinese University of Hong Kong

Assignment 2

Due date: 27 October 2023 (Fri) 23:59 Full mark: 100
Expected time spent: 3-5 hours

Aims: 1. Understand the knowledge about Support Vector Machine and Clustering.
2. Hands-on practice of the optimization problem in SVM.
3. Get familiar with how to use some tools (e.g., scikit-learn) to implement SVM.
4. Hands-on practice of implementation processes of K-means, DBSCAN and Hierarchical clustering.

Description:

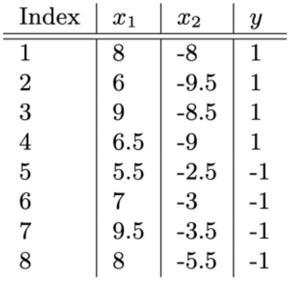
In Assignment 2, you will practice how to build a SVM classifier on a training set and evaluate it on a test set. Here, you will know how to use some toolboxes to implement SVM. You will practice on using concepts of K-means, DBSCAN and Hierarchical clustering to solve problems.

For some calculations, you can use the toolbox in Python or MATLAB or any other programming languages you are familiar with.

Questions:

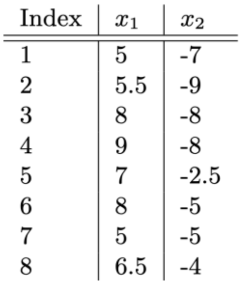
1. Consider the following training and test data set:

Training set



Index	x_1	x_2	y
1	8	-8	1
2	6	-9.5	1
3	9	-8.5	1
4	6.5	-9	1
5	5.5	-2.5	-1
6	7	-3	-1
7	9.5	-3.5	-1
8	8	-5.5	-1

test set



Index	x_1	x_2
1	5	-7
2	5.5	-9
3	8	-8
4	9	-8
5	7	-2.5
6	8	-5
7	5	-5
8	6.5	-4

(a) Set up the optimization problem using $(\alpha_1, \alpha_2, \dots, \alpha_8)$ and write down the dual problem of optimization. Then, given that the optimal $\alpha_1 = \alpha_8 = 0.32,\ \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7 = 0.$ Based on these $\boldsymbol{\alpha}$**,** which are the support vectors? Then, calculate the function of optimal hyperplane of this model. (10%)

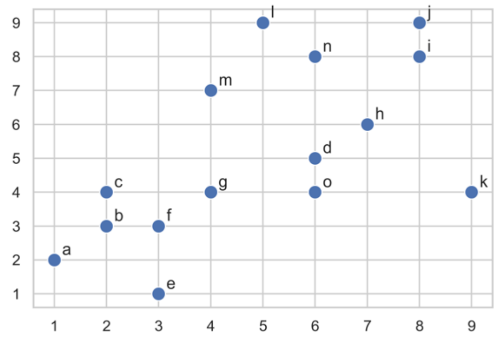
(b) Given the test data points, how can we use the optimal hyperplane to predict them? Please write corresponding formulas and then get the predictions. (10%)

(c) If we remove the 2nd data point in training set and use the remaining 7 points to train the SVM model, will the prediction of test data change? How about removing the 8th data point? (5%)

2. Use the training data attached in the blackboard to train a SVM model. Try to use what we have learned from the lecture (kernel function and soft-margin method) to optimize your model and make the accuracy of your model $\geq 86\%$ on the test set. Please try **at least two models** and each model should use different kernel function. You should give a brief introduction about your models (less than 50 words) and plot your decision boundaries with training set and test set in

2D figure (x−axis expresses x_1 while y−axis expresses x_2), respectively. Then, submit your code (Python, MATLAB or other programming languages you like). (25%)

3. Consider the following data and perform the K-means and DBSCAN algorithms using the Euclidean distance between points:



(a) Use K-means algorithm to perform three-class classification task. Assume that the initial cluster centers are (7, 4), (3, 6), (7,9) respectively. Calculate the updated cluster centers after the first iteration. (10%)

(b) Use DBSCAN to perform the clustering. Assume that $\epsilon = 2$ and *minpts*=2. List all core points, border point, noise point (exclude the point itself as its neighbor). (10%)

(c) Is *h* density reachable from *a*? Show the intermediate points on the chain or the point where the chain breaks of DBSCAN model. (5%)

(d) Show the density-based clusters of DBSCAN model. (5%)

4. Use the distance matrix in the following table to perform the hierarchical clustering with the **group average distance** and **max distance** respectively. Show your results by listing all intermediate updated table and drawing the final dendrogram. The dendrogram should clearly show the order in which the points are merged. (20%)

	A	B	C	D	E
A	0				
B	8	0			
C	6	7	0		
D	2	3	9	0	
E	1	6	5	4	0

Submission:

Submit a single file named <ID>_asmt2.pdf, where <ID> is your student ID.
Your file should contain the following header. Contact Professor Dou before submitting the assignment if you have anything unclear about the guidelines on academic honesty.

```
CSCI3230 / ESTR3108 2023-24 First Term Assignment 2

I declare that the assignment here submitted is original except for source
material explicitly acknowledged, and that the same or closely related material
has not been previously submitted for another course. I also acknowledge that I
am aware of University policy and regulations on honesty in academic work, and
of the disciplinary guidelines and procedures applicable to breaches of such
policy and regulations, as contained in the following websites.

University Guideline on Academic Honesty:
http://www.cuhk.edu.hk/policy/academichonesty/
Faculty of Engineering Guidelines to Academic Honesty:
http://www.erg.cuhk.edu.hk/erg-intra/upload/documents/ENGG_Discipline.pdf

Student Name: <fill in your name>
Student ID  : <fill in your ID>
```

Submit your files using the Blackboard online system.

Notes:

1. Remember to submit your assignment by 23:59pm of the due date. We may not accept late submissions.
2. If you submit multiple times, **ONLY** the content and time-stamp of the **latest** one would be considered.

University Guideline for Plagiarism

Please pay attention to the university policy and regulations on honesty in academic work, and the disciplinary guidelines and procedures applicable to breaches of such policy and regulations. Details can be found at http://www.cuhk.edu.hk/policy/academichonesty/. With each assignment, students will be required to submit a statement that they are aware of these policies, regulations, guidelines and procedures.