MATH 123 Math Aspects of Data Analysis - Spring 2023
Tufts University, Department of Mathematics
Instructor: James M. Murphy
Practice Midterm 2

Instructions: This exam has four questions and is out of a total of 100 points. Each question is worth 25 points. No graphing calculators, books, or notes are allowed. Be sure to show all work for all problems. No credit will be given for answers without work shown. If you do not have enough room in the space provided you may use additional paper. Be sure to clearly label each problem and attach them to the exam. You have 75 minutes. Good luck! :-)

Your Printed Name: Rubric

Problem	Score
1	
2	
3	
4	
Total	

Academic Honesty Certification:

I certify that I have taken this exam without the aid of unauthorized people or objects.

Signature: _____ Date: _____

QUESTION 1

(a) Describe, in detailed pseudocode, the DBSCAN clustering algorithm.

(b) Draw or describe in detail a data set on which DBSCAN can outperform K-means clustering. Explain.

(a.) Input: Data $\{x_i\}_{i=1}^n$

Parameters ε, MinPts

1) Label a point either a core, border, or noise point where

a.) x core if $|B_\varepsilon(x) \cap \{x_i\}_{i=1}^n| \geq$ MinPts

b.) x border if x not core but $\exists\, x_j$ s.t. $x_j \in B_\varepsilon(x)$ and x_j is core

c.) x is noise if it is neither core nor border.

2.) Say: x is directly density reachable from y if a.) y is core
b.) $x \in B_\varepsilon(y)$

- x is density reachable from y if $\exists$ path $x_0, x_1, \ldots, x_L$ s.t. $x_0 = x$, $x_L = y$ and x_i is directly density reachable from x_{i-1}, $i = 2, \ldots, L$
- x and y are density connected if there exists a core point z s.t. both x and y are density reachable from z

3.) Compute as clusters the maximal sets of density connected points

(b.) for example,

For some choice of ε, MinPts, the DBSCAN algorithm will learn the X and · clusters; K-means will always fail.

QUESTION 2

Let $L \in \mathbf{R}^{n \times n}$ be an (unnormalized) graph Laplacian corresponding to some weight matrix $W \in \mathbf{R}^{n \times n}$.

(a) Prove that L has at least one eigenvalue of 0.

(b) Describe in detail an example of a W matrix for which the corresponding L has exactly $k = 3$ eigenvalues of 0.

a.) Recall $L = D - W$, where $D_{ij} = \begin{cases} 0, & i \neq j \\ \sum_{k=1}^{n} W_{ik}, & i = j \end{cases}$

Then notice if $\mathbb{1} \in \mathbb{R}^{n \times 1}$ is the all 1s-vector,

$$\therefore L\mathbb{1} = D\mathbb{1} - W\mathbb{1}$$

$$= \left(\sum_{k=1}^{n} W_{1k}, \sum_{k=1}^{n} W_{2k}, \ldots, \sum_{k=1}^{n} W_{nk}\right) - \left(\sum_{k=1}^{n} W_{1k}, \ldots, \sum_{k=1}^{n} W_{nk}\right)$$

$$= (0, 0 \ldots, 0)$$

$$= 0 \cdot \mathbb{1},$$

i.e., $\mathbb{1}$ is an eigenvector of L with eigenvalue 0.

b.) $L = D - W$ where W corresponds to 1 graph with exactly 3 connected components, i.e.

} no edges between the 3 clusters

$$W = \begin{pmatrix} \mathbb{1}_{n/3 \times n/3} & 0 & \\ 0 & \mathbb{1}_{n/3 \times n/3} & \\ & & \mathbb{1}_{n/3 \times n/3} \end{pmatrix}$$

QUESTION 3

Let $\{x_i\}_{i=1}^n \subset \mathbf{R}^D$ be data with labels $\{y_i\}_{i=1}^n$, $y_i \in \{-1, 1\}$.

(a) Explain, in detailed pseudocode, the k-NN classifier.

(b) What is the complexity of labeling a new data point with this algorithm?

(c) Should this be considered an acceptable complexity or not? Discuss.

a.) 1.) Pick a #NN K.

2.) Compute the K-NN of a new observation x, and get their labels $\{y_j\}_{j=1}^K$

3.) Set the label of x to be the mode of $\{y_j\}_{j=1}^K$

b.) We need to compute NN, so $O(nD)$ unless fancy NN tricks are used

c.) For large n and D and with many points to test, probably not!

QUESTION 4

Let $\{x_i\}_{i=1}^n \subset \mathbb{R}^D$ be data with labels $\{y_i\}_{i=1}^n$, $y_i \in \{-1, 1\}$.

(a) Write the loss function for the hard margin support vector machine. Interpret both the function to be optimized and the constraints.

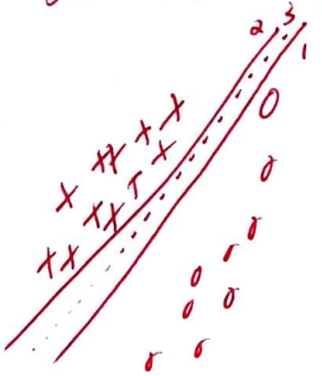
(b) Why is the idea of "maximizing the margin" relevant from the standpoint of labeling new data?

a.) Learn w^*, b^* s.t. $w \in \mathbb{R}^D$, $b \in \mathbb{R}$

margin maximizer: $\|w\|^2$ s.t. (class separability): $y_i(X_i w^T + b) \geq 1$ for all $i = 1, \ldots, n$

$(w^*, b^*) = \underset{(w,b)}{\arg\min}$

b.) Should (?) help us generalize well to new data:



line 3 (dotted) may generalize better than lines 1, 2?