MTH 4224/CSE 4224 Final Exam

Deadline: Thursday, May 2 by 11:59 PM (submit in Canvas) **Max Points:** 110/100

Instructions

- To receive full credit, **define what any variables you introduce represent** in words and **show all mathematical work**.
- **Solutions must be handwritten, either with a tablet/stylus or on paper.**
- You may use notes, books, or other sources, but you **may not** communicate with people about the exam (except for Dr. White).

Problems

1. What is the difference between bagged and boosted classifiers? (Max 2 sentences.) What are two machine learning problems boosting is intended to solve? [10]

2. Suppose we use a linear basis function expansion model for regression with basis functions $\{1, x, \sin x, \cos x\}$ with L^2. regularization. Write the corresponding least squares optimization problem and derive an exact matrix expression for the coefficients of the model. [15]

3. Explain how DBSCAN is trained using full mathematical details and write pseudocode for it. Give one advantage of DBSCAN over K-means clustering. [15]

4. Formulate the optimization problem corresponding to multidimensional scaling (MDS) mathematically. Describe geometrically what solving this problem does. (Max 1 sentence.) [10]

5. What a density attractor? What is the role of density attractors in DENCLUE? (Max 4 sentences.) [10]

6. Suppose you are given a task to cluster a dataset of Spotify users. You are given their top 20 songs labeled with genre (1 of 6 broad categories), duration, 10-second audio clip (stored as a long numerical time series), and features including mean loudness (from -60 to 0 dB), dancability (0 to 1), and energy (0 to 1).

   (a) How would you preprocess the data into a suitable data matrix? [10]

   (b) What is your approach to clustering? Include which methods would you try first, how you would progress, and how you would assess success practically. [10]

7. What is the purpose of principal components analysis (PCA)? (Max 2 sentences.) Formulate and solve the optimization problem to find the 1D PCA projection. [20]

8. What are three ways machine learning can be used in your area of study? (Max 3 sentences.) [5]