Dr. Christoph F. Eick

Review List Midterm2 Exam DS 1 COSC 3337

Tuesday, November 14, 2023 11:30a-12:45p

Last updated: November 3, 9p

Class room: SW 101

The exam will be “open books and notes” (but use of computers & internet is **not** allowed) and will center on the following topics (at least 85% of the questions will focus on material that was covered in the lecture); bringing calculators is okay. There will be no programming in this exam.

1. \*\*\*\* Neural networks (class transparencies, textbook pages 249-262); content of the  two introductory videos by 3blueonebrown about neural networks:
[Introduction to Neural Networks](https://www.bing.com/videos/search?q=neural+network+video&view=detail&mid=54402D363ABB8903202F54402D363ABB8903202F&FORM=VIRE) (watch the whole video)
[Weight Learning in Neural Networks](https://www.youtube.com/watch?v=IHZwWFHWa-w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=3&t=0s) (just watch the first 15 minutes of the second video)
2. \*\*\*\* Density Estimation (Lecture Slides which discuss naïve and parametric Density Estimation and non-parametric density estimation)

Also take a look of the first 3 pages of: [Multivariate kernel density estimation - Wikipedia](https://en.wikipedia.org/wiki/Multivariate_kernel_density_estimation)

1. \*\*\*\*\*\*\* Clustering (you should have in depth knowledge of K-Means, PAM/k-medoids, Hierarchical Clustering, DBSCAN; cluster evaluation, particularly silhouette and methods which use regression; all class transparencies which introduced these measures and algorithms are relevant)
2. \*\* Association Analysis centering on its goals and the APRIORI algorithm.
3. \*\*\* Data Storytelling (class transparencies; maybe a short essay-style question)
4. \*\* Outlier Detection (class transparencies centering on model-based, density-based and distance-based outlier detection); will only ask basic questions!
5. \* Model Evaluation (should know how N-fold cross validation work and what validation sets are used for!)

Other relevant material: Group Homework Credit Slides of Presentations that covered the above topics.

Relevant Lecture Slides:

IV Classification (,[Neural Networks](https://www2.cs.uh.edu/~ceick/UDM/NN2023.pptx),)

 V Density Estimation ([Naive and Parametric Density Estimation](https://www2.cs.uh.edu/~ceick/UDM/PDE.pptx) ([PDE Task](https://www2.cs.uh.edu/~ceick/UDM/GroupF-PDE-Task.pptx) (added on Nov. 6, 2023)), [Non-parametric Density Estimation](https://www2.cs.uh.edu/~ceick/UDM/NPDE.pptx)(slides have been added on November 3, 2023!)VI Clustering ~~and Similarity Assessment~~ **Similarity Assessment is not relevant for Midterm2!** ([Introduction](https://www2.cs.uh.edu/~ceick/UDM/DS1_clustering1.pptx), [Density-based Clustering Centering on DBSCAN](https://www2.cs.uh.edu/~ceick/UDM/dm_clustering2.pptx), [Hierarchical Clustering](https://www2.cs.uh.edu/~ceick/UDM/DS1_HC.pptx), [Cluster Validity](https://www2.cs.uh.edu/~ceick/UDM/DS1_CLVALIDITY.pptx), R-scripts demonstrating: [K-means/medoids](https://www2.cs.uh.edu/~ceick/UDM/r-clustering.r), [DBSCAN](https://www2.cs.uh.edu/~ceick/UDM/dbscan.r), [More on PAM and using PAM/DBSCAN with dist-objects (not relevant and covered in 2022)](https://www2.cs.uh.edu/~ceick/UDM/pam.r); [Clustering Exercises K-Means, HC, and DBSCAN](https://www2.cs.uh.edu/~ceick/UDM/Clustering_Exercises1.pdf))
VII [Outlier Detection](https://www2.cs.uh.edu/~ceick/UDM/udm_od22.pptx)
VIII Association Analysis: [Brief Introduction to Association Analysis Centering on APRIORI](https://www2.cs.uh.edu/~ceick/UDM/AA1.pptx)
IX [Data Storytelling](https://www2.cs.uh.edu/~ceick/UDM/DataStorytelling.pptx)