Review List Second Exam COSC 6335 Tu., Nov. 26, 2024, 2:30-3:45p

The exam will be open books and notes, but the use of computers is not allowed, but using calculators is okay. It will take 75 minutes and will count 25% towards your overall grade. The second exam will be a little different from Midterm Exam in that it will contain essay-style tasks and in that it asks questions about some reading material listed below.

1. \*\*\*\* Deep Learning Centering on Autoencoders:

VAE: [https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73](https://urldefense.com/v3/__https:/towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73__;!!LkSTlj0I!AzFFmh43oOfokMSm5bMfxED9-0Zr2pe_9MH0dOoymdRjvfqHlzaqP1s4OhYr6KeYQMVw0ZlT_kDA0sYcH6I$) (Excluding PCA )

[Deep Generative Learning](https://www.youtube.com/watch?v=3G5hWM6jqPk&list=PLtBw6njQRU-rwp5__7C0oIVt26ZgjG9NI&index=5) (watch the first 22 minutes of this video and 31:00-39) and Part 2:    [Review Neural Network Basics, Autoencoders, Language Models and Convolutional neural networks (CNN))](https://www2.cs.uh.edu/~ceick/DM/DL-Intro.pptx); taught by Mahin on October 31, 2024 Remark: Language Models and CNNs will not be covered in the exam!

2. \*\*\*\*\*\*\* Clustering: K-means, FCM, EM, Clique: Lecture slides, should be able to apply K-means, FCM to an example; should have knowledge about what mixtures of Gaussians are and basic ideas what Clique and EM do and how they work.

Read: [Gaussian Mixture Models — PyPR v0.1rc3 documentation (sourceforge.net)](https://pypr.sourceforge.net/mog.html)

Read: [Fuzzy clustering - Wikipedia](https://en.wikipedia.org/wiki/Fuzzy_clustering)

Clique: [Understanding CLIQUE Algorithm in Data Science. (janbasktraining.com)](https://www.janbasktraining.com/tutorials/clique-algorithm)

3. \*\*\* Association Analysis focusing on Association Rule Mining; should be able to apply the APRIORI algorithm, to an example. What is association analysis?

4. \*\*\*\*Decision Trees, and General Topics for Classification, particularly decision tree induction algorithm, overfitting, classification model performance evaluation (covered class transparencies)

6. \*\*\*\*\* Neural networks (class transparencies); 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)

7. \*\*\* Be able to answer questions about the ICDM 2022 Best Paper Award paper you reviewed in Task5!

8. \*\*\* Spatial Data Mining [Introduction to Spatial Data Mining](https://www2.cs.uh.edu/~ceick/DM/spdm_intro.pptx) [Spatial analysis - Wikipedia](https://en.wikipedia.org/wiki/Spatial_analysis) (only read the first 2 paragraphs and the sections titled Spatial Data Analysis, Spatial Autocorrelation and Heterogenity and Spatial Regression; there will not be any questions about other section of the Wikipedia Spatial Analysis document!

9. \*\*\*\*\* Write an essay centering on one theme of the two themes covered by the two GHC discussions; you will be given 2 essay topics to choose from!

Study the GHC slides which cover the above topics.