Review List Final Exam Data Mining Thursday, December 7, **2p**

The exam will be open books and notes, but the use of computers is not allowed, but using calculators is okay. It will be in the same room SEC 105 you had your midterm exam and starts at 11a. It will take about 105 minutes and counts 27% towards the course grade.

The final 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 and CNN:

VAE: [https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73](https://urldefense.com/v3/__https%3A/towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73__;!!LkSTlj0I!AzFFmh43oOfokMSm5bMfxED9-0Zr2pe_9MH0dOoymdRjvfqHlzaqP1s4OhYr6KeYQMVw0ZlT_kDA0sYcH6I$) (Excluding PCA )

CNN: [https://cs231n.github.io/convolutional-networks/?fbclid=IwAR3mPWaxIpos6lS3zDHUrL8C1h9ZrzBMUIk5J4PHRbKRfncqgUBYtJEKATA](https://urldefense.com/v3/__https%3A/cs231n.github.io/convolutional-networks/?fbclid=IwAR3mPWaxIpos6lS3zDHUrL8C1h9ZrzBMUIk5J4PHRbKRfncqgUBYtJEKATA__;!!LkSTlj0I!BifWKc2wsopUZ71PZAytecXyHUBDqbxBzvkm3ow5I_C9omLPNix2xy2XGGu6e6IsloOSmOT7XugllVzJkjA$)

[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);

**Remark:** Will only ask very basic questions about CNNs but deeper questions about auto-encoders; language models are not relevant for the final 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)

5. \*\*\*SVM (class transparencies, also scan through <http://en.wikipedia.org/wiki/Kernel_method>)

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. \*\* Data Storytelling  [Data Storytelling](https://www2.cs.uh.edu/~ceick/DM/DataStorytelling.pptx) slides

8. \*\*\* Parametric Density Estimation [Naive, Parametric](https://www2.cs.uh.edu/~ceick/DM/PDE.pptx) Density estimation slides

9. \*\*\* Similarity Assessment: Should be able to design a distance function based on a specification.

The final exam might contain 1-2 more essay-style questions. Study the GHC slides which cover the above topics.