Dr. Christoph F. Eick

Review List Final Exam Data Mining COSC 6335

Friday Dec. 9, **2p** in our class room.

*Last updated: December 6, 2022 at 5p*

The exam will be “open books and notes”; the use of computers is forbidding, but using calculators is okay!

1. \*\*\*\* Be able to answer questions concerning the content of the ICDM 2021 best paper award paper!
2. \*\*\*\*\* Clustering Algorithms; only the following algorithms are relevant
   1. FCM (Fuzzy C-means) lecture slides [Fuzzy clustering - Wikipedia](https://en.wikipedia.org/wiki/Fuzzy_clustering)
   2. EM lecture slides [Gaussian Mixture Models — PyPR v0.1rc3 documentation (sourceforge.net)](http://pypr.sourceforge.net/mog.html)
   3. DENCLUE (lecture slides, DENCLUE paper)
   4. K-means (lecture slides)
3. \*\*\*\*\*Decision Trees, and General Topics for Classification, particularly decision tree induction algorithm, overfitting, classification model performance evaluation (covered class transparencies)
4. \*\*\*SVM (class transparencies, <http://en.wikipedia.org/wiki/Kernel_method>)
5. \*\*\*\* Neural networks (class transparencie); 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)
6. \*\*\*\* Deep Learning: CNN[[1]](#footnote-1) (lecture slides & [CS231n Convolutional Neural Networks for Visual Recognition](https://cs231n.github.io/convolutional-networks/?fbclid=IwAR3mPWaxIpos6lS3zDHUrL8C1h9ZrzBMUIk5J4PHRbKRfncqgUBYtJEKATA)), and Autoencoders (lecture transparencies of lectures Nov. 4 and Dec. 2, [Variational autoencoder - Wikipedia](https://en.wikipedia.org/wiki/Variational_autoencoder))
7. \*\* Apriori Algorithm (lecture transparencies)
8. \*\*\* Non-parametric Density Estimation (lecture transparencies, [Kernel density estimation - Wikipedia](https://en.wikipedia.org/wiki/Kernel_density_estimation) )

All GHC credit slides which covered the above topics are relevant for the exam.

You should have detailed knowledge concerning the following algorithms, concepts and procedures: Apply non-parametric density estimation techniques to an example, APRIORI; K-Means, FCM (Fuzzy C-Means) and DENCLUE, weight computation and backpropagation in NNs; decision tree induction algorithm; information and GINI gain computations; Gaussian mixture models.

The final exam will counts approx. 26 % towards the overall course grade and should take about 100 minutes.

1. Will only ask very basic questions about CNNs! [↑](#footnote-ref-1)