Dr. Christoph F. Eick

Review List Final Exam Data Mining COSC 6335

Thursday, December 10, **2p** in Blackboard

*Last updated: December 7, 11p*

The exam will be “open books and notes”; there will be no R-programming in this exam:

1. \*\*\*Neural Networks, CNN, and Autoencoders (lecture transparencies, text book pages 249-262, <https://towardsdatascience.com/understanding-cnn-convolutional-neural-network-69fd626ee7d4> and <https://www.jeremyjordan.me/autoencoders/> )
2. \*\*\* Outlier Detection (lecture transparencies and textbook pages 703-732)
3. \*\*\* Sequence Mining (lecture transparencies and textbook pages 464-473)
4. \*\*\* Introduction to Density Estimation (lecture transparencies, <https://en.wikipedia.org/wiki/Multivariate_normal_distribution>, [Kernel density estimation - Wikipedia](https://en.wikipedia.org/wiki/Kernel_density_estimation) )
5. \*\*\* Similarity Assessment (Class Transparencies, capability to design a distance function for a dataset)
6. \*\*\*\*\*\*\*\*\*\*\* Clustering centering on objectives, cluster validity (only overview, correlation-based methods and silhouette are relevant for the exam), K-means, PAM, EM, DBSCAN, DENCLUE (lecture transparencies, DENCLUE 2.0 paper, [Gaussian Mixture Models — PyPR v0.1rc3 documentation (sourceforge.net)](http://pypr.sourceforge.net/mog.html), textbook pages 525-531, 534-542, 549-553, 565-574, 579-584)
7. \*\* Introduction to Spatial Data Mining (lecture transparencies and <https://en.wikipedia.org/wiki/Spatial_analysis> (discussions of Geospatial Analysis is not relevant for the exam! )

You should have detailed knowledge concerning the following algorithms, concepts and procedures: Apply density estimation techniques to an example, create a distance function for a dataset, a priori based sequence mining algorithm; K-Means, PAM, EM, DBSCAN and DENCLUE.

The final exam will counts 29% towards the overall course grade and should take about 100-105 minutes.