Homework 3

Unsupervised Learning • STAT GR 5244 • Fall 2025

Assigned: November 10

Due: November 24

Instructions:

- *You may work with others on this homework assignment but all solutions must be written up and submitted individually.*
- *All homework assignments must be submitted in pdf format and should be at most* **8 pages** *in length. Any material beyond 8 pages will not be graded.*
- *You must submit all code used to complete this homework assignment as an appendix (this can go beyond the 8 pages). Failure to submit code will result in 20% reduction.*
- *You are permitted 4 total late days on homework assignments throughout the semester. Any late days taken beyond these 4 will incur a 20% reduction per day.*
- *This homework assignment is worth* **40 points**. *But, there are up to* **60 points worth of extra credit available**.

1. (40 points + 10 points extra credit) Graphical Models. This problem asks you to download daily stock returns; there are many python packages that can help with this and `yfinance` is one such option.
   (a) Download and process daily stock return data from January 2021 - present for the following 15 companies: ["AAPL","MSFT","GOOGL","AMZN","META","NVDA","JPM","BAC","XOM", "CVX","JNJ","PFE","WMT","PG","KO"]. You may want to apply a log or other transform to this data. Briefly visually explore this data before applying graphical models.
   (b) Fit undirected graphical models via the Graphical Lasso and a non-parametric version of the Graphical Lasso. You should use proper hyperparameter tuning to determine the sparsity of the graph.
   (c) Fit a directed graphical model via the PC algorithm or a structural equation model. You should again tune any hyperparameters to determine graph sparsity.
   (d) Compare and interpret your graphical model estimates. Do the graphs relate companies as you would expect?
   (e) (10 points extra credit) Fit a Granger causal graphical model or a Vector Autoregressive Model (order 1 is fine) to this same data. Properly tune any hyperparameters. Does this order 1 model fit the data well? Interpret your results and compare these with your previous graph estimates.
2. (50 points extra credit) Density estimation & Generative Models. For this problem, you should use `sklearn`'s `digits` data.
   (a) (15 points) Apply kernel density estimation to this data and use this to generate new samples. You may apply kernel density estimation in a different space than the original data. You should also tune the bandwidth parameter. Are the samples generated from this density estimator recognizable as digits?
   (b) (15 points) Build a Variational Autoencoder (VAE) OR a Generative Adversarial Network (GAN) to generate new samples. Rigorously tune any hyperparameters. Are the samples generated recognizable as digits? Are all possible digits generated?
   (c) (20 points) Apply a denoising diffusion model to generate new samples from this data. Comapre the samples generated to those from the above methods and discuss your results.