C-M-009: Machine Learning Lab II '20-21

Lab. Project (Given December 30, 2020; Due January 5, 2021)

Your answers must be entered in Moodle by midnight of the day it is due. You must submit the code as well as a self-contained PDF which has the approach, an explanation of the implementation, the output as well as anything else asked by the question. Marks devoted to this Project are indicated in the "Syllabus" sheet that was provided the first day of class.

There are several days to submission. Please plan your work and submit before the deadline. Late submissions will not be accepted

Do not copy from someone else or discuss with anyone else. Doing so will result in an "F" for everyone involved – those who copied and those who allowed the copying to occur

---

1. A pharmaceutical company has hired you, a machine learning expert, to help them with chemoinformatics i.e. the generation of new chemical structures using machine learning. Assume chemical structures are represented using SMILES strings (see `https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system`) and you have access to structures that exist in the ZINC database (see, `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1360656/`).

   Ideally, new structures have to be screened but here we will only implement the part that is responsible for generating new structures. You are to implement an approach that will generate new molecules. Submit 3 new structures that you come up with using your implementation along with your code.

   Hint: VAE