C-M-009: Machine Learning Lab II '21-22

Lab. 4/Project (Given December 27, 2021; Due January 8, 2022)

Your answers must be entered in LMS 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 Lab. are indicated in the "Syllabus" sheet that was provided the first day of class.

---

1. A pharmaceutical company has hired you, a machine learning expert, to help them with chemoinformatics i.e. the generation of new chemical structures efficiently. Create a system that you can deliver to the pharmaceutical company based on your knowledge of machine learning that meets the objective of the pharmaceutical company.

   Hint: One approach is to use SMILES strings (see `https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system`) to represent molecules and the ZINC database (see `http://zinc.docking.org`) of commercially available stable molecules. There is even a library that fetches molecules in the ZINC database in the SMILES format (see `https://github.com/rasbt/smilite`). **(100 points - 40% of the overall grade)**