Computational Chemistry 2,   CHEM-E4225,   Exercise 6   2.12.2022

In this exercise you need **python which contain the sklearn library**. It can be started with 'module load python'   all .py programs can be run with command python prog.py

More information of the sklearn manual pages. Search sklearn.ensambe for RandomForrest (Warning these are very technical)

There are a lot of example file in /home/kari/CC2-2022-examples

To see what is in this dir type   ls -l /home/kari/CC2-2022-examples   (ls is the list command)

you can copy the example files to your own directory: cp /home/kari/CC2-2022-examples/h2o.inp .
(there is a dot at the end it is your working directory)

1) Use the UMAP-numbers.ipynb   program to analyse the handwritten number recognition. It will have two data set data_id=41082 and mnist_784 they are rather large so the ML will take a bit of time. See the scattering of the numbers, like 3 and 8. which of the data set are cleaner.

To run the program you type jupyter notebook   UMAP-numbers.ipynb
you need jupyter notebook and UMAP   install them with pip install notebook   pip install umap-learn

2) Use the PCA-numbers.ipynb   program to analyse the PCA effect of numbers. Use 5, 15 and 25 components in the PCA analysis. See the singular values and the reconstructed numbers fit.

3) Use PCA-claisen.ipynb to analyse the molecular clustering. You need the claisen-smiles-cyclic.csv data. The program is not very nice. Look what is SMILES representation of molecules. Try to understand the fingerprint data (NOT how they have been done). Use the UMAP method: variable cPCA = 'UMAP' and PCA method   cPCA = 'PCA' (in the second box). There are also the 3D plots but in notebook you cannot rotate them. (or I don't know how to able that).

you need rdkit   install it with pip install rdkit

The sklearn library can be started with 'module load python'

To see geometries you can use ase, module load ase, ase-gui …

The instructions of mylly2 are included.

In the first time make your own directory in /home/kari/CC2-2022-results
mkdir /home/kari/CC2-2022-results/ossi   (ossi should be your own name)
At end of exercise copy the results to your result dir: cp *out /home/kari/CC2-2022-results/ossi