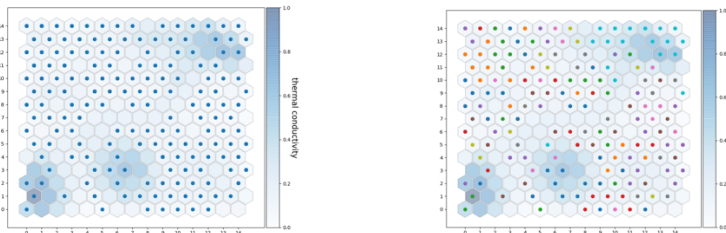


Homework4

Q1: Self-organizing map analysis of thermal conductivity dataset

The following code SOM-HexagonalTopology.py apply SOM unsupervised clustering to our thermal conductivity dataset and shows the cluster map, and maps the data samples to each of the clusters with colors showing its 1 out of 10 grades corresponding to their percentiles.

- (1) Add your code from line 99, so that your SOM map can show different colors for samples of different thermal conductivity grade (from 0 to 9, corresponding to their percentile).



Hint: this function from scipy which can calculate the percentile of a value in a list of numbers

from scipy import stats

print(stats.percentileofscore(target, 500))

You need to install the minisom by: pip3 install minisom

<https://github.com/JustGlowing/minisom> more info.

- (2) If possible, try to fix the legend bar so that the color show the range of thermal conductivity values. (optional, bonus points: 10)

Q2: Genetic programming for symbolic regression

Study this fastsr symbolic regression package

<https://github.com/cfusting/fast-symbolic-regression>

read the thermal_dataset.csv file, use all the numeric columns except the y-exp and y-theory columns as the X_train, use the y-exp as the y_train
 train a symbolic regression model for this dataset
 print out the final regression score
 print out the formula of the best individual
 plot the final regression scatter plot.