Implementation note 02272024

Tuesday, February 27, 2024 10:12 AM

When splitting train and test, pay attention to data imbalance if combine different groups

Remove datapoints where UP_SEC_HMW is greater than 100

Scale y and scale back? Robust transform

Note the only buffer is categorical, other data is actually ordinal

Complete the modality similarity analysis and find out similar modalities and combine them to conduct analysis

python - Optimize the Kernel parameters of RBF kernel for GPR in scikit-learn using internally supported optimizers - Stack Overflow

Understanding Kernels in Gaussian Processes Regression — Prog-ML (code-first-ml.github.io)
Variance parameter in the optimized kernel makes the difference when scale y

MyGPR2 outperforms default
GaussianProcessRegressor in uncertainty

```
my_gpr = MyGPR2(kernel=kernel,
n_restarts_optimizer=9)
    #my_gpr =
GaussianProcessRegressor(kernel=kernel,
n_restarts_optimizer=9)
```

https://github.com/scikit-learn/scikit-learn/blob/main/sklearn/gaussian_process/_gpr.py

References

.. [RW2006] `Carl E. Rasmussen and Christopher K.I. Williams,
"Gaussian Processes for Machine Learning",
MIT Press 2006 <https://www.gaussianprocess.org/gpml/chapters/RW.pdf>`_