|  |  |  |
| --- | --- | --- |
| **Fall 2024** | **Report #5 – 10/01/2024** | **Dingyi Nie** |

**——————————————————————————————————————**

**I. Task Achieved Last Week**

**——————————————————————————————————————**

* Simple feature analysis on P12

**——————————————————————————————————————**

**II. Feedback and Interaction**

**——————————————————————————————————————**

* **Prof. Kuo’s Feedback**
  + Last week Prof. Kuo suggested we change direction and explore a non-imputing approach. Specifically, we ignore the temporal pattern and simply extract means and deviations for each feature (channel)’s original signal and differenced signal.

**——————————————————————————————————————**

**III. Report**

**——————————————————————————————————————**

For the whole dataset array shaped (N, L, C), for each channel in range(0, C-1), we capture 4 features:

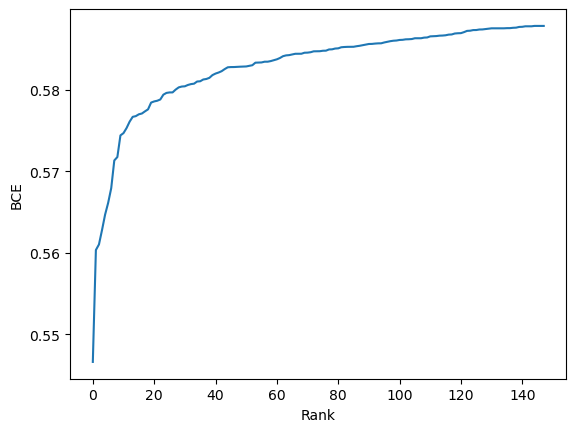
* Average of observed values
* Standard deviation of observed values
* Average of differenced observations (delta signal)
* Standard deviation of differenced observations (delta signal)

This will result in a feature array of size (N, 4C). For P12 dataset, C = 37.

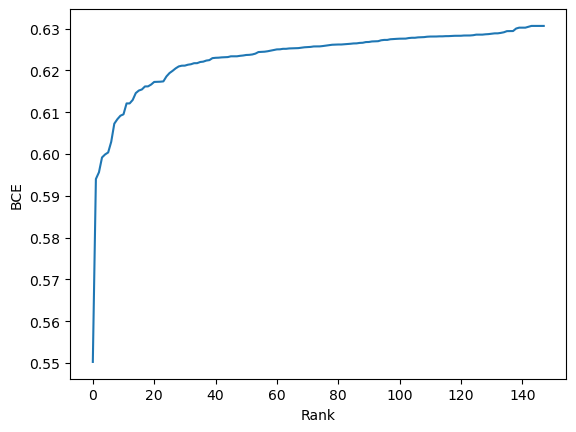
We split P12 into 8:1:1 train vs valid vs test sets. If in a specific channel of an episode there are no observed value, use the global average on train set to impute the average observation. For the other 3 features, impute with 0.

By fitting an XGBoost classifier directly to the (N, 4C) training feature array, we get a 86.55% test AUROC with the best parameters searched (n\_estimators=100, max\_depth=5, learning\_rate=0.1).

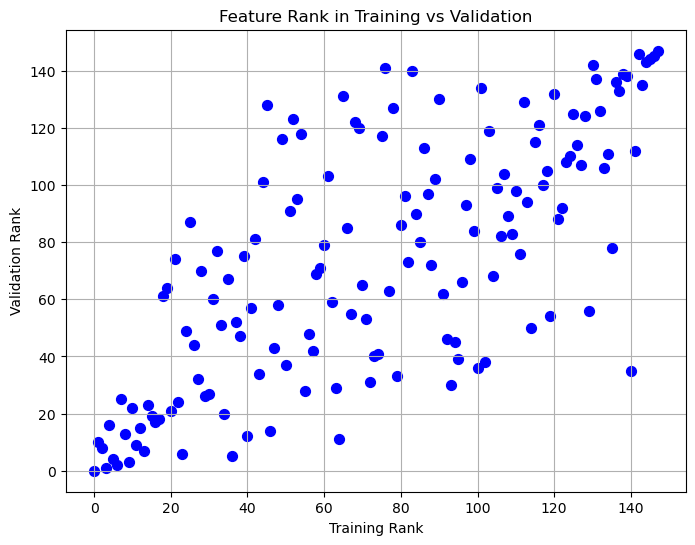
DFT rank elbow curve on test set:



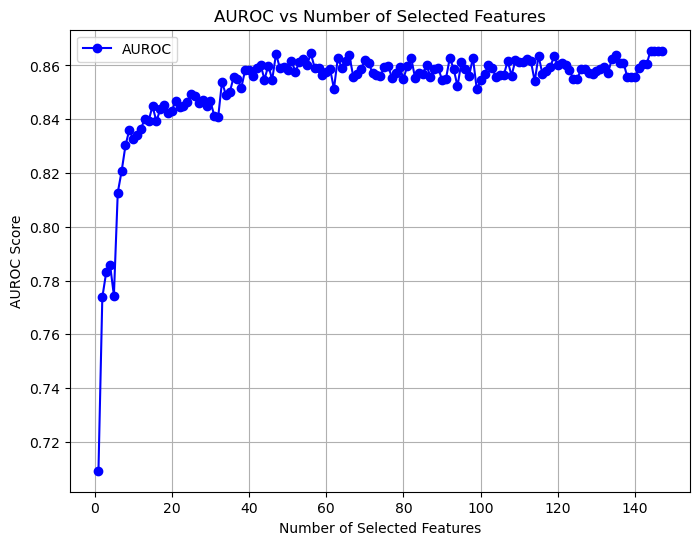
DFT elbow curve on valid set:



Joint feature rank scatter plot:



Select top-k ranked features and fit a classifier again on selected features, we can plot a curve of test AUROC vs number of selected features:



This shows that DFT has limited improvement on P12 classification score. With n\_selected = 80, we get a test AUROC of 86.62%.

**——————————————————————————————————————**

**IV. Next Steps**

——————————————————————————————————————

* Apply same analysis on more medical tabular datasets?

**——————————————————————————————————————**

**V. Milestone**

——————————————————————————————————————

* Codes are available here: <https://github.com/d9sus4/GL-TS>