

Updates for Week 5

This week was mostly a development week — didn't get a lot of time to play with the systems but I tried to add as many useful things as possible (for analysis next week).

Important Updates:

- 1) Minimum Sample Size
- 2) Massaging the feature vector more
- 3) Recursive Labelling (Ran it with the entire set)
- 4) Regular Labelling (Ran it with the entire set)
- 5) Threshold Labelling (Coded it, but I haven't run it yet)
- 6) Self Organizing Map Adaptation (given online code — fixed bugs plus extra plots)
- 7) QQ plots
- 8) Probability Plots
- 9) Correlation Matrix (Pearson, Kendall, and Spearman)
- 10) Linear Discriminant Analysis Plotting, and (LDA + PCA plots)
- 11) Recursive Feature Extraction (RFE) (Determine number of features with lots of cross validation)

New Files:

- 1) QQ_plots....ipynb
- 2) Visualizing Features Chosen.ipynb
- 3) actually_label_data.ipynb
- 4) pickle files
- 5) min_sample_size.ipynb
- 6) label_data.py
- 7) SOMPY/* (all files)

Things To Do For Next Week:

- 1) Actually use the correlation matrix and LDA to analysis features
- 2) Tune the SOM to optimize
- 3) Cross validation stuff
- 4) Try it with threshold labelling

New Diagrams:



