Region-Detect

region_detect is a python script originally part of a DMR detection pipeline. Given a csv file of methylation calls across multiple samples for individual CpG sites, the script will attempt to aggregate neighboring CpG sites into regions. The script uses a sliding window approach to aggregate CpG methylation values based on proximity to the next CpG site and the maximum delta methylation between the aggregated region and the next CpG site.

minimum requirements:

1. python 2.7.11 - https://www.python.org/downloads/release/python-2711/
2. pandas - https://github.com/pydata/pandas.git
3. scipy - https://sourceforge.net/projects/scipy/files/scipy/
4. numpy - https://sourceforge.net/projects/numpy/files/NumPy/

0) Input methylation table ./test_data/meth_table.csv

The methylation table should be in csv file format with headers. The first column should be labeled 'POS' for the CpG coordinates. Expected coordinate values include the choromosome label with the chromosome position separated by an underscore. All subsequent columns contain the methylation call for a sample. The methylation call should range from 0.0-1.0. 'NA' should be used in place of any missing values.

1) Aggregate neighboring CpGs into fragments region_detect.py

- [in] -dmr_join - path to methylation table
- [in] -dmr_bpdist - max distance to neighboring site for aggregation in bp (default=500)
- [in] -dmr_cpgmin - minimum number of sites for aggregation (default=2)
- [in] -dmr_maxdelta - maximum differnce in methylation between neighbors to aggreagte

```
cd <region_detect path>
python region_detect.py -dmr_join ./test_data/meth_table.csv -dmr_bpdist 500 -dmr_cpg
min 2 -dmr_maxdelta 0.15
```