

ECE: 523 HW #5

David Akre

Semi-Supervised Learning: Self-Training

1) An Experiment on Synthetic Data

The console output for (1):

Part 1 - Error of self training algorithm with 10% of the data labeled

Error at the first pass through - 0.35

Error half way through - 0.37

Error at the end - 0.26

Part 2 - Error of self training algorithm with 25% of the data labeled

Error at the first pass through - 0.60

Error half way through - 0.57

Error at the end - 0.43

2) An Experiment on Real World Data

1) Brief discussion on whether semi-supervised learning helped on real-world data sets:

It appears that semi-supervised learning did help overall in most samples (e.g. 70% of the data sets for the semi-supervised learning algorithm had a greater than 50% accuracy: annealing_test, breast-cancer-wisc-diag_test, credit-approval_test, breast-cancer-wisc_test, conn-bench-sonar-mines-rocks_test, cylinder-bands_test, and bank_test). The other 30% of the data samples had around 40% accuracy which is less than 50%; in other words the semi-supervised learning was unsuccessful at predicting the correct values in these cases. In summary, it seems that the semi-supervised self-training learning algorithm did a good job in terms of minimizing error and offering valid predictions on most of the data samples used (e.g. 70% of the time). It is important to note that there is some randomness on selecting which parts of the data set get selected for the supervised and unsupervised parts of the algorithm so the results may differ on from execution to execution runs.

The console output for (2):

```
*****
File: annealing_test.csv
Supervised Learning Accuracy: 1.00
Unsupervised Learning Accuracy: 0.72
Total Semi Supervised Learning Accuracy: 0.86
*****
File: breast-cancer-wisc-diag_test.csv
Supervised Learning Accuracy: 0.86
Unsupervised Learning Accuracy: 0.97
Total Semi Supervised Learning Accuracy: 0.91
*****
File: congressional-voting_test.csv
Supervised Learning Accuracy: 0.20
Unsupervised Learning Accuracy: 0.60
Total Semi Supervised Learning Accuracy: 0.40
*****
File: cylinder-bands_test.csv
Supervised Learning Accuracy: 0.67
Unsupervised Learning Accuracy: 0.62
Total Semi Supervised Learning Accuracy: 0.64
*****
File: conn-bench-sonar-mines-rocks_test.csv
Supervised Learning Accuracy: 0.67
Unsupervised Learning Accuracy: 0.67
Total Semi Supervised Learning Accuracy: 0.67
*****
File: credit-approval_test.csv
Supervised Learning Accuracy: 1.00
Unsupervised Learning Accuracy: 0.87
Total Semi Supervised Learning Accuracy: 0.94
*****
File: bank_test.csv
Supervised Learning Accuracy: 0.48
Unsupervised Learning Accuracy: 0.87
Total Semi Supervised Learning Accuracy: 0.67
*****
File: heart-hungarian_test.csv
Supervised Learning Accuracy: 0.33
Unsupervised Learning Accuracy: 0.65
Total Semi Supervised Learning Accuracy: 0.49
*****
File: breast-cancer_test.csv
Supervised Learning Accuracy: 0.25
Unsupervised Learning Accuracy: 0.62
Total Semi Supervised Learning Accuracy: 0.44
```

File: breast-cancer-wisc_test.csv

Supervised Learning Accuracy: 0.88

Unsupervised Learning Accuracy: 0.93

Total Semi Supervised Learning Accuracy: 0.90