

CS 557 STATISTICAL PATTERN RECOGNITION AND LEARNING
FALL 2016
ASSIGNMENT 5

DUE: November 06, 2016.
PROBLEM

Use assignment 4 dataset and apply SOM. The kaggle website is:

<https://inclass.kaggle.com/c/scrambled-ocr>

Initialize the weight vectors randomly and run the algorithm for around 10-50 iterations. Repeat SOM with different combinations of the following parameters:

- grid size: 10x10, 20x20, 30x30
- decay rate: 2 different values. You can use the same value for both lambda and beta. One good value is .0001
- radius for neighborhood function: 2 different values. You can use $\sigma_0 = 2$ (optional)

NOTE: It may take time to run SOM on a bigger grid so don't leave the simulation results for the last minute.

NOTE:

- When you make the figures, make sure you make a legend and label all axis. You can use the plot command to make a scatter plot. For example `plot(x,y,'*g')` will make a scatter plot in green color using the symbol '*'.
- Use `pdist2`, `eigs`, sort routines of Matlab. AVOID THE USE OF LOOPS

TO SUBMIT

1. Softcopy of all code and report on slate
2. **Hard copy** of a report which is **not more than two pages** long that describes all the results of your experiments AND YOUR CONCLUSION and COMMENTS ON THE RESULTS. In the results you have to include the results on the training set and also your public leaderboard scores for the test set.