**Lab 6 – Unsupervised Learning**

1. What is the purpose of the neighborhood function in the SOM? How does it change the learning?
2. A simplistic intruder detection system for a computer network consists of an attempt to categorize users according to (i) the time of day they log in, (ii) the length of time they log in for, (iii) the types of programs they run while logged in, (iv) the number of programs they run while logged in. Suggest how you would train a SOM classifier to perform the categorization. What preprocessing of the data would you do, how much data would you need, and how large would you make the SOM? Do you think that such a system would work for intruder detection?
3. A bank wants to detect fraudulent credit card transactions. They have data for lots and lots of transactions (each transaction is an amount of money, a shop, and the time and date) and some information about when credit cards were stolen, and the transactions that were performed on the stolen card. Describe how you could use a competitive learning method to cluster people's transactions together to identify patterns, so that stolen cards can be detected as changes in pattern. How well do you think this would work? There is much more data of transactions when cards are not stolen, compared to stolen transactions. How does this affect the learning, and what can you do about it?
4. Consider the following data set consisting of the scores of two variables on each of seven individuals (subjects). This data set is to be grouped into two clusters using k-Means clustering algorithm.



Initialization - .  As a first step in finding a sensible initial partition, let the A & B values of the two individuals furthest apart (using the Euclidean distance measure), define the initial cluster means, giving:



Complete the iterations needed to finalize your answers (maximum 5 iterations will suffice).