2413, Machine Learning, Homework 4 Solutions
Due Date: 27/11/2013
Universität Bern

Question 1 (1 points) Implement the k-means clustering algorithm, and apply it to the two data sets attached to this document. The function specification is the following:

```
[C, A] = Kmeans_YourName(X, Cinit)
% input: X: the data matrix, each column represents a data point
%        Cinit: the initial cluster centers, each column represents
%               a cluster center
% output: C: the cluster centers after the k-means, the format is
%            the same as Cinit
%         A: row vector of labels, A(i) = j means that X(:,i)
%            belongs to the cluster j, ie. C(:,j) is the closest
%            cluster center
%
% Your code here
```

You have to substitute "YourName" with your name and surname in capital letters. Add comments in your code to explain the most important parts of your algorithm.

Question 2 (2 points) Run the k-means algorithm several times on the first data set with different initialisation. Do you always get the same solution? If not, why?

Question 3 (2 points) You can see, that the k-means algorithm does not cluster the second dataset according to the ground truth. Why does this happen?
Hint: The k-means algorithm is an optimisation method, it minimises a cost function. Calculate the cost function for the ground truth, and check, if it is optimal or not.