Machine Learning
Exercise Sheet 6

Prof. Dr. Dr. Lars Schmidt-Thieme, Martin Wistuba
Information Systems and Machine Learning Lab
University of Hildesheim

November 25th, 2014
Submission until December 2nd, 13.00 to wistuba@ismll.de

Exercise 19: (5 Points)

a) Explain the optimization function of the K-Means clustering in your own words.

b) What is the primary difference between K-Means and K-Medoids?

c) Apply the K-Means Clustering for two iterations on the data in Table 1 for $k = 4$. The first cluster center was randomly chosen to be $(3, 3)$.

x	y	x	y
-4	4	3	3
-4	3	4	3
-3	3	4	4
-4	-3	3	-3
-4	-4	4	-4
-3	-3	4	-3
3	3	4	-4

Exercise 20: (5 Points)

a) What do Gaussian Mixture Models and K-Means have in common? What are the differences.

b) In the lecture it was mentioned that K-Means are a special case of Gaussian Mixture Models. Explain.

Exercise 21: Agglomerative Hierarchical Clustering (5 points)

Use single-link, complete-link and average-link agglomerative hierarchical clustering to cluster the points P1 to P8 using the given distance matrix. Draw the dendogram.

	P1	P2	P3	P4	P5	P6	P7	P8
P1	0	5	6	3.61	7.07	7.21	8.06	2.24
P2		0	6.08	4.24	5	4.12	3.16	4.47
P3			0	5	1.41	1.41	7.28	6.40
P4				0	3.61	4.12	7.21	1.41
P5					0	1.41	6.71	5
P6						0	5.39	5.39
P7							0	7.62
P8								0

Exercise 22: Frequent Itemset Mining (5 points)

a) What is the pruning criteria of the apriori breadth-first search algorithm with respect to the support values of supersets?

b) Consider the transactions dataset shown in Table 1, which is recorded in an imaginary local store, which sells only four products, namely bread, beer, bratwurst and borecole. Find the frequent item sets having a minimum support value of three, by using the NAIVE breadth-first search techniques.

Customer ID	Product 1	Product 2	Product 3	Product 4	Transaction Time
Customer 1	Bratwurst	Bread	Beer	Borecole	25/11/14 8:21:12
Customer 2	Beer	Borecole			25/11/14 8:43:09
Customer 3	Bratwurst	Bread			25/11/14 9:15:12
Customer 4	Borecole	Bratwurst	Beer		25/11/14 9:23:53
Customer 5	Bread	Bratwurst	Beer		25/11/14 10:00:01
Customer 6	Bratwurst	Bread			25/11/14 10:12:21
Customer 7	Beer	Bratwurst			25/11/14 10:45:54
Customer 8	Bratwurst	Borecole			25/11/14 10:59:32

c) Repeat the computations of b) with the pruning trick of a). How many support calculations do you need now?