Final examination
Machine Learning and Data Mining

Lessons notes and slides are permitted.

NB : Presentation will be taken into account when given a mark. Marks are also subject to modification.

Exercise 1 : Classification **(7 pts)**

1) Suppose that we want to select between two prediction models, M1 and M2. We have performed 10 rounds of 10-fold cross-validation on each model, where the same data partitioning in round i is used for both M1 and M2. The error rates obtained for M1 are 30.5, 32.2, 20.7, 20.6, 31.0, 41.0, 27.7, 26.0, 21.5, 26.0. The error rates for M2 are 22.4, 14.5, 22.4, 19.6, 20.7, 20.4, 22.1, 19.4, 16.2, 35.0. Comment on whether one model is significantly better than the other considering a significance level of 1%.
2) Compare classification and prediction tasks.
3) Is an ensemble method always better than a single classifier ? Explain

Exercise 2 : Itemset and Sequence Mining **(13 pts)**

id	time	items
s_1	10	A,B
	20	B
	30	A,B
	40	A,C
s_2	20	A,C
	30	A,B,C
	50	B
s_3	10	A
	30	B
	40	A
	50	C
	60	B
s_4	30	A,B
	40	A
	50	B
	60	C

Consider the database table above. Each sequence is comprised of itemset-events that happen at the same time. For example, sequence s_1 can be considered to be a sequence of itemsets $(AB)_{10}(B)_{20}(AB)_{30}(AC)_{40}$, where symbols within brackets are considered to co-occur at the same time, which is given in the subscripts. The itemsets-sequences can be of any length as long as they are frequent. The minsup is set to 3.

X is a frequent maximal pattern in a data set S if there exists no frequent proper super-pattern Y and X satisfies minimum support.

X is a closed pattern in a data set S if there exists no proper super-pattern Y such that Y has the same support count as X in S, and X satisfies minimum support.

1. Given *minsup = 3*, are the following sequences frequent ? (2 pts)
   a. without time constraints
      i- (AB)B ii- (AC)A iii- AC iv- ABC
   b. with time-constraints **Max-Gap = 10**
2. Find all frequent itemset-sequences *without time constraints*. Provide details. (5 pts)
3. Give two frequent maximal itemset-sequences. Explain. (1 pt)
4. Give two frequent closed itemset-sequences that are not maximal. (1 pt)
5. **ITEMSET** mining :
   a. Explain the similarity and difference between positive and negative borders of frequent itemsets. (1 pt)
   b. Given a negative border of frequent itemsets, describe an algorithm that can generate the positive border of these frequent itemsets. (3 pts)