RMIT University
School of Science
COSC2110/COSC2111 Data Mining
Tutorial Problems Week 11

1. Using J48 as the base classifier show the operation of bagging on the following data. Assume there will be 3 classifiers in the ensemble.

Run	Supervisor	Operator	Machine	Overtime	output
1	Patrick	Joe	a	no	high
2	Patrick	Sam	b	yes	low
3	Thomas	Jim	b	no	low
4	Patrick	Jim	b	no	high
5	Sally	Joe	c	no	high
6	Thomas	Sam	c	no	low
7	Thomas	Joe	c	no	low
8	Patrick	Jim	a	yes	low
9	Patrick	Sam	c	no	???

2. Show the operation of Adaboost for $k = 2$ on the above data.

3. A random forest classifier uses decision trees in which the split attribute is chosen at random. Show the how a random forest of 3 trees could be generated and used on the above data.

4. Boosting requires a classifier in which the instances can be weighted. How can this be achieved without modifying the base classifier to use weighted instances?