1.

Association rules are designed to find the positive co-relation between the item sets based on the frequency the items appear together. However sometimes the associated items are negatively co-related for example if the total number of transactions is 100,

-10 of those contain both milk and toothpaste

-70 transactions have milk in them but not toothpaste and

-4 have just toothpaste.

Thus the confidence will be 10/(10+4)= 0.7 which gives us a high confidence value however in general these products are weakly associated to each other hence this confidence value cannot be trusted. To overcome this drawback we use another measure known as LIFT.

Lift enables us to lift up \*rise\* the probability of presence of {y} when {x} is present upon the probability of having {y} in the itemset without the presence of {x}

Lift =( transactions X and Y) /(Transactions of x)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(transactions containing y)

So the probability of Milk in the cart is (70+10) 80/100=0.8 when we have no knowledge about the presence of toothpaste. These calculations indicate that the presence of toothpaste has negative co-relation with the milk and the probability of the milk has been reduced to 0.7.

If we calculate the co-relation using Lift:

07/0.8 = 0.87 which is < 1 , this shows us that presence of toothpaste on the itemset does not positively influence the chances of milk occurring in the dataset.

2.

If we run the Apriori algorithm using weather.nominal.arff in Weka we can the see the output contains minimum support and minimum confidence value which is user specified(default) and the number of cycles performed which is 17. What this means is that the cycle starts at upperBoundMinSupport which is 100% (represented as 1 in weka) decreasing it each time by 5%(0.05) delta value and stops the cycle when the user defined numRules(10) is attained or at predifned lowerBoundMinSupport (default 10%). It tells us the size of Large Item sets with support >=2 (0.15 x14 instances =2)in case of weather.nominal.arff it is one Itemset with a size 12, 2 itmesets with size 47, 3 itemsets with size 39 and 4 itemsets with size 6. Also at the end of the report we can find the best rules which have maximum support and fulfills the minimum confidence criteria .

Leverage is the measure of difference between the probability of x and y appearing together in a dataset with the probability of x and y being statistically dependent to each other. Leverage = P (xUy)- P(x).P(y). Setting up leverage threshold would enable user to find the frequency of items appearing in the dataset with the user defined minimum support value and then filter these itemsets based on leverage value. Range of Leverage is [-1,1] and a value of 0 means independence. Conviction is the measurement of probability of x occurring without y when they are dependent on each other to the definite probability of x occurring without y in the given dataset. Conviction= P(x)P(not y)/P(x U not y). Conviction ranges from [0, infinity].

The result output also gives the value of Lift. In a way Conviction and Lift are similar except the fact that conviction also considers the absence of the value of ‘y’ and has monotones(unchanging) for confidence and lift values. Lift is the measurement of occurrence of x and y together in a dataset to the value of them being independent of each other statistically . Lift overcomes the drawback of Leverage as it is able to calculate the values for lower probable itemsets known as ‘rare itemsets’ .

Lift = P(xUy)/P(x).P(y) or Lift=confidence/support

An example of Lift is explained in the previous answer and thus we know that a value of Lift less than 1 would indicate a negative dependence of x and y and a value larger than 1 would indicate a positive relation of x and y. Thus instead of just depending on the value of confidence we can know for sure using lift if the relationship between x and y is positive or negative. Range of lift is between [0, infinity] and a value of 1 is considered the ideal value meaning independence between x and y.

3.

By default the algorithm mines using Nominal association rule which has no particular end target. It locates all the possible associations in the itemsets. When you set car=true in Apriori algorithm class association rules are mined instead of the general association rules. However in cases the user is interested in finding a particular association among the itemsets. For example let us consider a supermarket transaction dataset which has ‘n’ number of transactions and each transaction has a class label of ‘y’.

‘T’ = set of all items in the dataset

‘Y’ = set of all class labels

T intersection Y = phi

Thus X 🡪 y, assuming X ⊆ *T*, and *y* ∈ *Y*. Enabling CAR would help us find the ruleitems that consist of the user defined minimum support value with a class label of Y. The mining is done in just one step as we have already defined a target label.