CSCI B 565: Assignment #4

Due on Saturday, April 9, 2016

Prof. Predrag

Vivek Patani

Contents

Problem 1

Listing 1 shows a Perl script.

Listing 1: Sample Perl Script With Highlighting

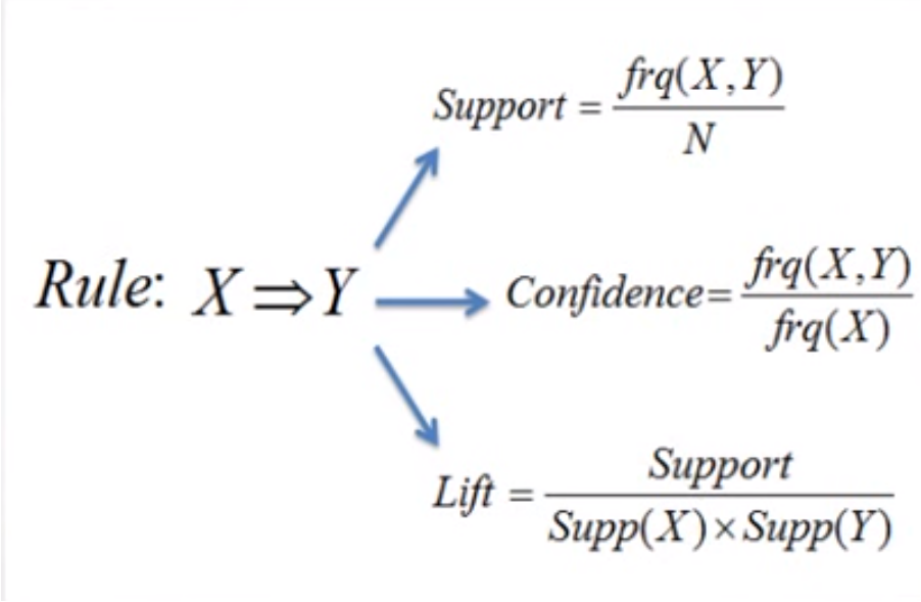
```
#!/usr/bin/perl

use strict;
use warnings;

for (1..99) { print $_." Luftballons\n"; }

# This is a commented line

my $string = "Hello World!";

print $string."\n\n";

$string =~ s/Hello/Goodbye/;

print $string."\n\n";

test();

exit;

sub test { print "All good.\n"; }
```

Problem 2

Problem 3

What are **Association Rules**?

- The idea of Association Rules is that it can be defined as a pattern stating the occurrence of one event, by which another event occurs with a certain probability.
- In other words, they are simple If/Else statements which help us discover patterns between unrelated data.
- The interesting part of Association Rules is that they help us learn relationships of objects which are frequently collectively used.
- The goal is to find all sets (only important ones) having $support > minsup$ (minimum support).
- Followed by checking which rules cross a certain level of confidence. It means that not all rules generated are important, instead simply pick those rules which have $confidence > minconf$.
- The two most basic and important things to look out while mining Association Rules are **Support** & **Confidence**. Also **Lift** is another alternative, due to the shortcomings of the prior.



- Some places where this is used is:
  - Market Basket Analysis.
  - Medical Diagnosis
  - Protein Sequences
  - Census Data

The basic **Apriori Algorithm** is divided into two parts:

- Finding Frequent Itemsets.
  - Generating Candidate Itemsets.
  - Filtering the useful candidates.

Problem 3 continued on next page...

- Mining/Discovering Unknown Patterns from frequent itemsets.

1. How to find the Frequent Itemsets?
    - The very basic need is to specify a **Support level**.
    - Next step consists of actually scanning through the input transaction set and only collecting that cross the threshold of Support. In other words, eliminate those who do not matter and contribute towards forming interesting patterns.
    - Once we filter based on the support level, we start building candidate sets.
    -

References

- https://www.researchgate.net/publication/238525379_Association_rule_mining-_Applications_in_various_areas
- https://www.youtube.com/watch?v=RHkvnRemaLE&index=3&list=PLVOXKA8fjRuvTVJt_n1rkRl6n3sGcyY6a
- http://aimotion.blogspot.com/2013/01/machine-learning-and-data-mining.html
- http://www2.ift.ulaval.ca/~chaib/IFT-4102-7025/public_html/Fichiers/Machine_Learning_in_Action.pdf