

Data Mining - The Diary

Rodion “rodde” Efremov

March 17, 2015

1 Introduction

This document is my learning diary written on behalf of Data Mining course led at spring term 2015 at University of Helsinki.

2 Week 1

The **support count** $\sigma(X)$ of an item set X is the amount of transactions containing X ($X \subset t_i$). Basically, we were computing support counts for various itemsets with the exception of applying additional constraints to the queries (such as particular grade range).

The **support** of an item set X is $\sigma(X)/N$, where N is the amount of all transactions. Support of X may be thought of as a classical probability of a random transaction containing X .

An **association rule** is an implication of the form $X \rightarrow Y$, where X and Y are itemsets having no items in common. The interpretation of an association rule is that if a transaction contains X , it “tends” to contain Y as well. Note that “tends” depends on parameters we specify to a data mining system. **Support** of an association rule $X \rightarrow Y$ is

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}.$$

Support of the rule R may be thought of as a classical probability of R appearing in a random transaction. **Rule confidence** gives the probability of Y appearing in the same transactions with set X and is defined as

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}.$$