# Week 4 Tutorial exercises (March 31st 2006)
# Sequential Pattern Analysis

***Exercise 1.*** ***AprioriAll***

Apply the AprioriAll algorithm to the following customer sequence dataset using minimum support s=33%. Identify the maximal sequence patterns.

| *S.ID* | *Sequence* |
|---|---|
| *1* | *<{1 5}{2}{3}{4}>* |
| *2* | *<{1}{3}{4}{3 5}>* |
| *3* | *<{1}{2}{3}{4}>* |
| *4* | *<{1}{3}{5}>* |
| *5* | *<{4}{5}>* |

***Solution:***

Find the large 1-sequences

| Sequence | Support |
|---|---|
| <1> | 4 |
| <2> | 2 |
| <3> | 4 |
| <4> | 4 |
| <5> | 4 |

Find the large 2-sequences

| Sequence | Support |
|---|---|
| <1,2> | 2 |
| <1,3> | 4 |
| <1,4> | 3 |
| <1,5> | 2 |
| <2,3> | 2 |
| <2,4> | 2 |
| ~~<2,5>~~ | 0 |
| <3,4> | 3 |
| <3,5> | 2 |
| <4,5> | 2 |

Find the large 3-sequences

| Sequence | Support |
|---|---|
| <1,2,3> | 2 |
| <1,2,4> | 2 |
| <1,3,4> | 3 |
| <1,3,5> | 2 |
| ~~<1,4,5>~~ | 1 |
| <2,3,4> | 2 |
| ~~<2,3,5>~~ | 0 |
| ~~<2,4,5>~~ | 0 |
| ~~<3,4,5>~~ | 1 |

Find the large 4-sequences

| Sequence | Support |
|---|---|
| <1,2,3,4> | 2 |

The maximal sequences:
<1,2,3,4> is a maximal sequence. The only Large 3-sequence not contained in <1,2,3,4> is <1,3,5>.
The only Large 2-sequence neither contained in <1,2,3,4> or <1,3,5> is <4,5>.
Thus the maximal sequences are : <1,2,3,4> , <1,3,5> and <4,5>.

***Exercise 2. GSP***

Apply the GSP algorithm to the following dataset using minimum support s=3 transactions. Show the candidates and the resulting large sequential items.

| SID | Sequence |
|---|---|
| 10 | <a(ac)(adc)> |
| 20 | <(ba)(fb)a> |
| 30 | <(ab)bfb(ae)> |
| 40 | <a(af)d> |
| 50 | <d(fac) > |
| 60 | <(adf)(ae)> |

***Solution:***

Scan 1:

| Candidate | Support |
|---|---|
| **a** | **6** |
| ~~**b**~~ | ~~**2**~~ |
| ~~**c**~~ | ~~**2**~~ |
| **d** | **4** |
| ~~**e**~~ | ~~**2**~~ |
| **f** | **5** |

**<a> <d> <f>**

Scan 2:

| | **<a>** | **<d>** | **<f>** |
|---|---|---|---|
| **<a>** | **<aa>:5** | ~~**<ad>:2**~~ | **<af>:3** |
| **<d>** | ~~**<da>:2**~~ | ~~**<dd>:0**~~ | ~~**<df>:1**~~ |
| **<f>** | **<fa>:3** | ~~**<fd>:1**~~ | ~~**<ff>:0**~~ |

| | **<a>** | **<d>** | **<f>** |
|---|---|---|---|
| **<a>** | | ~~**<(ad)>:2**~~ | **<(af)>:3** |
| **<d>** | | | ~~**<(df)>:1**~~ |
| **<f>** | | | |

**<aa> <af> <fa> <(af)>**

***Exercise 3.* FreeSpan**

Apply FreeSpan to the previous sequence database.

***Solution:***

| Candidate | Support |
|---|---|
| a | 6 |
| ~~b~~ | ~~2~~ |
| ~~e~~ | ~~2~~ |
| d | 4 |
| ~~e~~ | ~~2~~ |
| f | 5 |

**F_list= <a>:6 <f>:5 <d>:4**

Project over <a>, <f>, and <d>

<a> projected database:

| SID | Sequence |
|---|---|
| 10 | <aaa> |
| 20 | <aa> |
| 30 | <aa> |
| 40 | <aa> |
| 50 | <a > |
| 60 | <aa> |

Frequent 2-sequences wrt <a>:

**<aa>:5**

<f> projected database:

| SID | Sequence |
|---|---|
| 10 | <aaa> |
| 20 | <afa> |
| 30 | <afa> |
| 40 | <a(af)> |
| 50 | <(af)> |
| 60 | <(af)a> |

Frequent 2-sequences wrt <f>:

**<af>:3**

**<fa>:3**

**(af):3**

<d> projected databases:

| SID | Sequence |
|---|---|
| 10 | <aa(ad)> |
| 20 | <afa> |
| 30 | <afa> |
| 40 | <a(af)d> |
| 50 | <d(af) > |
| 60 | <(adf)a> |

Frequent 2-sequences wrt <d>:

~~<ad>:2~~ ~~<fd>:1~~

~~<da>:2~~ ~~<df>:1~~

~~(ad):2~~ ~~(df):1~~

***Exercise 4.* PrefixSpan**

Apply PrefixSpan to the previous sequence database.

***Solution:***

| Candidate | Support |
|---|---|
| a | 6 |
| ~~b~~ | ~~2~~ |
| ~~e~~ | ~~2~~ |
| d | 4 |
| ~~e~~ | ~~2~~ |
| f | 5 |

PrefixSpan(<>,0,S) outputs:

**<a>:6 <d>:4 <f>:5**

Remove all non frequent items

Call PrefixSpan(<a>,1, $\mathbf{S}|_{<a>}$)

PrefixSpan(<d>,1, $\mathbf{S}|_{<d>}$)

PrefixSpan(<f>,1, $\mathbf{S}|_{<f>}$)

| $S\|_{<a>}$ |
|---|
| <a(ad)> |
| <fa> |
| <fa> |
| <(af)d> |
| <(_f)> |
| <(_df)a> |

Frequent elements:

**<a>:5 ➔ <aa>:5**

~~(_d):1~~ ~~<d>:2~~

**<f>:3 ➔ <af>:3**

**(_f):3 ➔ (af):3**

| $S\|_{<d>}$ |
|---|
| <> |
| < > |
| < > |
| < > |
| <(af) > |
| <(_f)a> |

Frequent elements:

~~<a>:2~~ ~~(_f):1~~

~~(af):1~~ ~~<f>:1~~

| $S\|_{<f>}$ |
|---|
| <> |
| <a> |
| <a> |
| <d> |
| < > |
| <a> |

Frequent elements:

**<a>:3 ➔ <fa>:3**

~~<d>:1~~