ממן 13/אלירן סבג

שאלה 1:

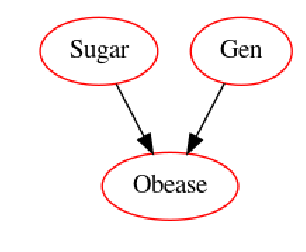
נתון:

P(S)=0.3, P(G)=0.4

נבנה את טבלת ההסתברויות עבור O.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S ^ ¬ G ¬ | S ^ G ¬ | S ^ ¬G | S ^ G |  |
| 0.2 | 0.6 | 0.4 | 0.8 | O |
| 0.8 | 0.4 | 0.6 | 0.2 | O ¬ |

על פי המודל הנתון אין צורך בנתונים יתר לחץ דם וסוכרת ( על פי כיוון החץ) , ולכן מודל מנוון יראה כך:



א. חשב P(O).

על פי השרטוט אין תלות בין S ל- G . נפתח את הביטוי לקבלת ההסתברות להשמנה (O)P.

P(O) = P(O|SG) \* P(S) \* P(G) + P(O|S¬G) \* P(S) \* P(¬G) +

P(O|¬SG) \* P(¬S) \* P(G) + P(O|¬S¬G) \* P(¬S) \* P(¬G) =

0.8\*0.3\*0.4+0.4\*0.3\*0.6+0.6\*0.7\*0.4+0.2\*0.7\*0.6 = 0.42

P(O) = 0.42

ב. חשב P(O|G).

**לפי נוסחת הההסתברות השלמה**

P(O|G) = P(O|SG) \* P(S) + P(O|¬SG) \* P(¬S) = 0.8\*0.3+0.6\*0.7 = 0.66

שאלה 2.

פתרון השאלה נמצא בקובץ Maman12\_Q2.ipynb המצורף.

שאלה 3:

‫

|  |  |
| --- | --- |
| מס' הזמנה | מספר פריטים |
| 1001 | ‪‪‬ ‫{i1,i4,i5} |
| 1024 | ‪‫}‪{i1,i2,i3,i5 |
| 1012 | ‫}‪{i1,i2,i4,i5 |
| 1031 | ‫}‪{i1,i3,i4,i5 |
| 1015 | ‫}‪{i2,i3,i5 |
| 1022 | ‫}‪{i2,i4,i5 |
| 1029 | ‫}‪{i3,i4 |
| 1040 | ‫}‪{i1,i2,i3 |
| 1033 | ‫}‪{i1,i4,i5 |
| 1038 | ‫}‪{i1,i2,i5 |

‫בהנחה‪:‬‬

‫‪Min\_support‬‬ ‫‪=60%‬‬

‫‪Min\_confidence‬‬ ‫=‬ ‫‪80%‬‬

‬‬

נבנה עץ על פי שלבי אלגוריתם FP-Groth :

ראשית נבנה טבלה ראשית שתספור כל פריט בודד, נמיין אותה בסדר יורד.

|  |  |
| --- | --- |
| Frequency | Item |
| 8 | i5 |
| 7 | i1 |
| 6 | i2 |
| 6 | i4 |
| 5 | i3 |

אנו צריכים תמיכה מינימאלית של 60% ( 6 מתוך 10 רשומות ) ולכן הפריט i3 יסונן.

הרשימה שקיבלנו לפי סדר יורד:

{i5:8, i1:7, i2:6, i4:6}

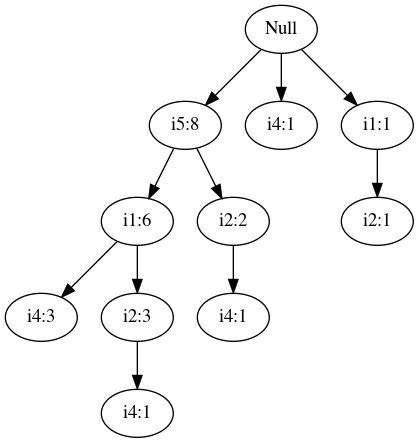
נרשום שוב את הטבלה אבל **הפריטים יסודרו בסדר יורד**. בעמודה השמאלית נבנה את העץ.

שורש העץ הוא NULL, לאחר מכן, לכל רשומה נעדכן את העלים בעץ ואת מספר ההופעות. במידת הצורך נפתח ענף חדש.

|  |  |  |  |
| --- | --- | --- | --- |
| מס' הזמנה | מספר פריטים | הסבר | העץ |
| 1001 | ‪‪‬{ i5, i1, i4 } | הרשומה החדשה בונה את העץ. הצומת הראשון מעליה יהיה תמיד:  Null.  כל פריט נספר פעם אחת. |  |
| 1024 | ‪‫{‪{ i5, i1, i2 | בהכנסת הרשומה השנייה, נמצאים כבר בעץ הפריטים 5 ו – 1 ולכן רק נוסיף למונה שלהם 1. הפריט 2 נוסף לראשונה ולכן ייצור עלה חדש עם המונה 1. |  |
| 1012 | ‫{ ‪{ i5, i1, i2, i4 | בהכנסת הרשומה השלישית, נמצאים כבר בעץ הפריטים 5 ,1 ו-2. ולכן רק נוסיף למונה שלהם 1. הפריט 4 נוסף לראשונה כקשק עם פריט 2 ולכן ייצור עלה חדש עם המונה 1. |  |
| 1031 |  | { i5, i1, i4 } |  |
| 1015 | { i5, i2 } | המשך בניית העץ נמצא בקובץ Q3\_fp\_grafh.ipynb |  |
| 1022 | { i5, i2, i4 } |  |  |
| 1029 | ‫{ i4 } |  |  |
| 1040 | ‪{ i1, i2 } |  |  |
| 1033 | { ‫{ i5, i1, i4 |  |  |
| 1038 | ‪{ i5, i1,i2 } |  |  |

בהתאם להסבר של הרשומות 1001, 1024, 1012 , 1031 נמשיך לבנות את כל העץ.

העץ המלא בסוף הבנייה (המשך בניית העץ נמצא בקובץ Q3\_fp\_grafh.ipynb):



על פי רמת הביטחון הדרושה, 80% לא התקבלו חוקי הקשר חזקים.

הערה: ניתן למשל לצורך הלימוד לבחור תנאים פחות נוקשים, למשל תמיכה של 0.3.

ואז הקבוצות שיתקבלו יהיו:

+---------+----+

| items|freq|

+---------+----+

| [1]| 7|

| [1, 5]| 5|

|[1, 5, 4]| 4|

| [1, 4]| 5|

| [3]| 5|

| [3, 1]| 3|

| [3, 5]| 3|

| [3, 2]| 3|

| [5]| 7|

| [5, 4]| 5|

| [2]| 6|

| [2, 1]| 4|

| [2, 5]| 4|

| [2, 4]| 3|

| [4]| 7|

+---------+----+

וחוקי ההקשר שיתקבלו:

+----------+----------+----------+-----------------+

|antecedent|consequent|confidence| lift|

+----------+----------+----------+-----------------+

| [1, 4]| [5]| 0.8|1.142857142857143|

| [5, 4]| [1]| 0.8|1.142857142857143|

| [1, 5]| [4]| 0.8|1.142857142857143|

+----------+----------+----------+-----------------+