WSI 21Z

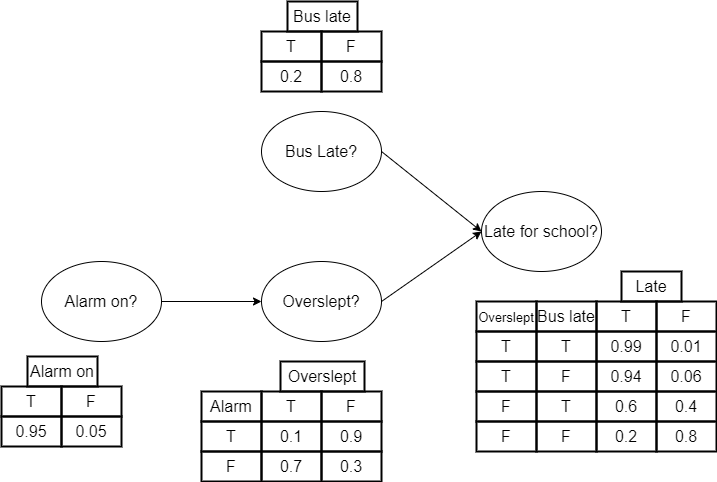
Laboratorium 7

Wojciech Gierulski

Napisać losowy generator danych działający zgodnie z rozkładem reprezentowanym przez daną sieć Bayes'a. Sieć powinna opisywać zależności między zmiennymi losowymi (binarne: 0/1) Na wejściu dany jest opis grafu połączeń/zależności między zmiennymi oraz tabele prawdopodobieństw warunkowych. Format danych wejściowych jest dowolny, może być i z pliku .py, ale nie jako MyCustomFancyGraphClassThatIWillForgetHowToUseSoon(1, a, 2, 84)... tylko raczej {'a': 'b', ...}, {'a': [0, 1], 'b': [[0,1],[1,0]]...} Wygenerować X próbek np. w liczbie [100, 1000, 10000] i sprawdzić czy są zgodne z rozkładem. Problem należy samemu zdefiniować, ale powinien zawierać kilka zmiennych i co najmniej jedną podwójną zależność.

## Założenia wstępne

Zdefiniowano problem prawdopodobieństwa spóźnienia do szkoły, który reprezentuje poniższa sieć Bayesa.



Rys. 1 – sieć Bayesa reprezentująca problem spóźnienia do szkoły

## Wyniki eksperymentów

Metodą propagacji przez sieć wygenerowano 100, 1000 i 10000 próbek losowych postaci:

{„Alarm on”: bool, ”Overslept”: bool, ”Bus late”: bool, ”Late for school”: bool}.

Następnie korzystając z tablic prawdopodobieństw ze zdefiniowanej sieci obliczono prawdopodobieństwa wystąpienia każdej możliwej próbki (czyli 16 prawdopodobieństw, bo próbka ma 4 zmienne). Wyniki eksperymentów i obliczone prawdopodobieństwa przedstawiono w tabeli 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nr próbki | % wystąpień w 100 próbkach | % wystąpień w 1000 próbkach | % wystąpień w 10000 próbkach | Prawdopodobieństwa wystąpienia [%] |
| 1 | 0.01 | 0.021 | 0.0174 | 0.01881 |
| 2 | 0.00 | 0.000 | 0.0000 | 0.00019 |
| 3 | 0.10 | 0.103 | 0.1021 | 0.10260 |
| 4 | 0.11 | 0.075 | 0.0657 | 0.06840 |
| 5 | 0.02 | 0.004 | 0.0073 | 0.00693 |
| 6 | 0.00 | 0.000 | 0.0002 | 0.00007 |
| 7 | 0.00 | 0.001 | 0.002 | 0.00180 |
| 8 | 0.01 | 0.001 | 0.0013 | 0.00120 |
| 9 | 0.06 | 0.076 | 0.0727 | 0.07144 |
| 10 | 0.00 | 0.005 | 0.0046 | 0.00456 |
| 11 | 0.12 | 0.137 | 0.1408 | 0.13680 |
| 12 | 0.55 | 0.539 | 0.5470 | 0.54720 |
| 13 | 0.01 | 0.027 | 0.0264 | 0.02632 |
| 14 | 0.00 | 0.003 | 0.0019 | 0.00168 |
| 15 | 0.00 | 0.002 | 0.0021 | 0.00240 |
| 16 | 0.01 | 0.006 | 0.0085 | 0.00960 |

Tabela 1

Z tabeli wynika, że generator próbek został zaimplementowany poprawnie. Procentowe wystąpienia danej próbki są zbliżone do prawdopodobieństwa wylosowania danej próbki. Wraz ze wzrostem wygenerowanych próbek te dwie wartości stają się coraz bliższe sobie.