

## Wprowadzenie do sztucznej inteligencji - ćwiczenie 7

### Zadanie

Proszę zaimplementować losowy generator danych, który działa zgodnie z rozkładem reprezentowanym przez daną sieci bayesowską. Sieć ta opisuje zależności między (zero-jedynkowymi) zmiennymi losowymi i dana jest w postaci opisu grafu połączeń oraz tabel prawdopodobieństw warunkowych. Wejście algorytmu: ile przykładów wygenerować, opis struktury prostej sieci (według własnego formatu) oraz tabele prawdopodobieństw należy wczytać z pliku tekstowego. Wyjście: plik tekstowy z przykładami. Strukturę sieci i tabele prawdopodobieństw widać na rysunku. Klasa to „Ache” (czy bolą plecy), pozostałe węzły to atrybuty („Back” to uszkodzenie kręgosłupa (drobne, czasem nie skutkujące bólem)). Wytworzony zbiór podzielić i użyć do treningu i testowania klasyfikatora utworzonego na wcześniejszych ćwiczeniach. Jakiego uzyskujemy wyniki? Wnioski?

### Rozwiązanie

Zdecydowałem się na przechowywanie struktury sieci w pliku json. Zaimplementowany przeze mnie algorytm pozwala na wczytanie dowolnych struktur w tym formacie, jeśli zależności zostaną zapisane w odpowiedniej kolejności (węzeł grafu od którego zależny jest inny, musi zostać zapisany wcześniej). Plik wczytywany jest do struktury opartej na Nodach, z których każdy przechowuje swoją nazwę, listę swoich zależności oraz słownik z prawdopodobieństwami.

### Wyniki i wnioski

Wygenerowałem 10000 próbek danych, a następnie użyłem algorytmu ID3 do predykcji (podział na zbiór treningowy i testowy w stosunku 3:2). Osiągnięty wynik był bardzo dobry - otrzymana trafność przewidywań to około 86%. Pobieżnie analizując tabele prawdopodobieństw dla zadania, można dojść do wniosku, że powinniśmy spodziewać się mniej więcej takiego wyniku, co oznacza, że algorytm wygenerował dane prawidłowo.