

WSI Ćwiczenie 7 – Sieci Bayesa

Kamil Marszałek 331401

Zadanie polega na stworzeniu generatora danych zgodnie z rozkładem reprezentowanym przez daną sieć Bayesa

Założenia

Plik, który opisuje sieć Bayesa musi mieć określoną strukturę:

- pierwsza linijka: NODES: {lista wierzchołków}
- druga linijka: EDGES: {lista krawędzi o budowie: A->B, gdzie A jest rodzicem B}
- w następnych linijkach: wartości prawdopodobieństw, tylko dla wartości prawdziwych

Dla zadanej sieci:

```
NODES: Chair, Sport, Back, Ache
EDGES: Chair->Back, Sport->Back, Back->Ache
P(Chair=true) = 0.8
P(Sport=true) = 0.02
P(Back=true|Chair=true,Sport=true) = 0.9
P(Back=true|Chair=true,Sport=false) = 0.2
P(Back=true|Chair=false,Sport=true) = 0.9
P(Back=true|Chair=false,Sport=false) = 0.01
P(Ache=true|Back=true) = 0.7
P(Ache=true|Back=false) = 0.1
```

Sprawdzenie poprawności napisanego generatora

Na podstawie częstości wystąpień dla dużej liczby próbek, widzimy, że generator działa poprawnie:

```
Test Results:
Node: Chair
  Condition: Chair=true, Observed: 0.8038, Expected: 0.8000
Node: Sport
  Condition: Sport=true, Observed: 0.0208, Expected: 0.0200
Node: Back
  Condition: Back=true|Chair=true,Sport=true, Observed: 0.9292, Expected: 0.9000
  Condition: Back=true|Chair=true,Sport=false, Observed: 0.1992, Expected: 0.2000
  Condition: Back=true|Chair=false,Sport=true, Observed: 0.9011, Expected: 0.9000
  Condition: Back=true|Chair=false,Sport=false, Observed: 0.0107, Expected: 0.0100
Node: Ache
  Condition: Ache=true|Back=true, Observed: 0.7028, Expected: 0.7000
  Condition: Ache=true|Back=false, Observed: 0.1004, Expected: 0.1000
```

Zastosowanie generatora z klasyfikatorem

Eksperymenty zostały przeprowadzone dla 10000 wygenerowanych danych. 60% wygenerowanych danych zostało użyte do trenowania klasyfikatora id3, natomiast 40% zostaje wykorzystana do testowania klasyfikatora. Przeprowadzono 100 eksperymentów. Dokładność minimalna, średnia, maksymalna oraz odchylenie standardowe zostały zapisane w tabeli poniżej:

Minimalna dokładność	Średnia dokładność	Maksymalna dokładność	Odchylenie standardowe
0,85125	0,8621	0,8725	0,0044

W tabeli poniżej przedstawiam jeszcze uśrednioną macierz pomyłek dla testowanego klasyfikatora, gdzie wartość pozytywna oznacza brak bólu pleców:

Prawdziwie pozytywne	Prawdziwie negatywne	Fałszywie pozytywne	Fałszywie negatywne
2960,38	488,19	332,79	218,64

Wnioski

Klasyfikator ID3 wykazuje wysoką skuteczność w oparciu o dane wygenerowane przez sieć Bayesa. Średnia dokładność wynosząca 86,21% oraz niskie odchylenie standardowe (0,44%) wskazują na stabilność i powtarzalność wyników w różnych próbach.

Poprawność działania generatora została potwierdzona przez porównanie wygenerowanych częstości z teoretycznym rozkładem prawdopodobieństw. Dzięki temu można wnioskować, że generator wiernie odtwarza zależności reprezentowane przez sieć.

Niska liczba fałszywie pozytywnych (332,79) oraz fałszywie negatywnych (218,64) wyników świadczy o dobrej zdolności modelu do klasyfikowania przykładów testowych. Wyniki te sugerują, że dane wygenerowane przez sieć Bayesa są dobrze zrównoważone i odzwierciedlają rzeczywiste zależności, co umożliwia efektywne uczenie się klasyfikatora.

Generowanie danych za pomocą sieci Bayesa może znaleźć zastosowanie w sytuacjach, gdzie brakuje wystarczających danych rzeczywistych do trenowania modeli.