

PS 9: Wykorzystanie danych w modelach decyzyjnych

Praca w grupach dwu- lub trzy-osobowych.

Zadanie 1 Uczenie parametrów sieci bayesowskiej z danych

Naucz się parametrów modelu *Credit.xdsl* z załączonych zbiorów danych: *credit10.csv*, *credit100.csv*, *credit1000.csv* i *credit10000.csv* o wielkości odpowiednio 10, 100, 1000 i 10000 rekordów używając opcji *Uniformize*. Przetestuj każdy z wynikowych modeli (*Validate* w menu *Learning*) na zbiorze *credit10000.csv* (opcja *Test Only*). Co zauważasz? Następnie powtórz ten eksperyment używając opcji *Keep original* z parametrem *Confidence* równym 1 i równym 100. Za każdym razem rozpoczynaj od oryginalnego modelu. Co zauważasz teraz?

Zadanie 2 Uczenie ciągłych rozkładów prawdopodobieństwa z danych

Dla zbioru danych *retention.txt*, naucz się rozkładu ciągłego prawdopodobieństwa zmiennej *salar* używając:

- (a) Rozkładu *Metalog()*.
- (b) Rozkładu *CustomPDF()*.

W obu wypadkach wybierz rozkład, który najlepiej pasuje do danych, używając różnych liczb kwantyli w (a) i przedziałów histogramu w (b). Stwórz prosty model sieci bayesowskiej składający się z dwóch węzłów *salar1* i *salar2*, w którym *salar1* zdefiniowany jest nauczonym rozkładem *Metalog()*, a *salar2* zdefiniowany jest nauczonym rozkładem *CustomPDF()*. Zmień liczbę próbek z domniemanej wartości 10000 do miliona i porównaj wygenerowane rozkłady *salar1* i *salar2*.

Po zakończeniu, wyślij krótki raport z pracy wraz z listą członków grupy na adres m.druzdzal@pb.edu.pl.