

Statystyka II. Kolokwium komputerowe. 16.06.2005

1. Wczytaj z pliku `...pokar/StatystykaMat/similMat.dat` macierz podobieństwa między aminokwasami (ściślej: wektorami własności aminokwasów) $C = (c_{ij}), 1 \leq i, j \leq 20$. Przekształć C na macierz odległości $D = (d_{ij})$ daną wzorem $d_{ij} = \sqrt{c_{ii} + c_{jj} - 2c_{ij}}$. Wykorzystaj macierz D do klasteryzacji aglomeracyjnej aminokwasów metodą *average linkage* (konieczna jest konwersja klasy *matrix* na *dist*). Wynik klasteryzacji, w postaci dendrogramu, zapisz do pliku postscriptowego.

2. Wskaźnik ubóstwa zdefiniowany w normie Eurostatu to prawdopodobieństwo, że dochód jest nie większy niż $0.6 * med$, gdzie *med* oznacza medianę dochodu. Załóżmy, że dochód ma rozkład *gamma(kształt = 4, skala = 1)*. Wygeneruj 1000-elementową próbę prostą z tego rozkładu i policz próbkowy wskaźnik ubóstwa. *Wskazówka* Kwantyle próbkowe liczby *quantile*. Policz teoretyczny wskaźnik ubóstwa.

3. Wczytaj dane demograficzne o szkołach holenderskich:

```
library(nlme); data(MathAchSchool); x=MathAchSchool[,-c(1,6)] .
```

Wyznacz, za pomocą *kde2d*, dwuwymiarową gęstość (*PRACAD, DISCLIM*) w dwóch grupach: dla szkół katolickich i publicznych. Policz wspólną informację $M(PRACAD, DISCLIM)$. Narysuj w jednym okienku, różnymi kolorami 2 wykresy konturowe tych gęstości.

4. Porównaj na danych z poprzedniego zadania nieparametryczne metody klasyfikacji zaimplementowane w funkcjach *knn* i *nnet* z bibliotek *class* i *nnet*. Zakładamy, że cecha przewidywana

```
y=ifelse(x[,2]=='Catholic',1,0) .
```

Wykonaj weryfikację krzyżową (4/5 danych do trenowania) i policz prawdopodobieństwo poprawnej predykcji (wektor 5-elementowy) dla obu metod.

5*. Napisz funkcję losującą prostą próbę z wielowymiarowego rozkładu normalnego $rnormMV(n, m, S)$, gdzie n jest liczbą niezależnych losowań, m – wektorem wartości oczekiwanych oraz S – macierzą kowariancji. *Wskazówka* Wykorzystaj *eigen* do przekształcenia wyniku *rnorm*.