

## Statystyka II. Kolokwium komputerowe 27.07.2005

1. Wczytaj z pliku `.../AAindex.dat` macierz danych zawierającą własności fizykochemiczne aminokwasów. Przedstaw te dane na płaszczyźnie dwóch pierwszych składowych głównych i znajdź numer odstającej obserwacji (będą dodatkowe punkty za wykorzystanie `eigen` zamiast `prcomp` lub `princomp`). Następnie wykonaj klasteryzację aglomeracyjną za pomocą funkcji `agnes` z biblioteki `cluster`. Dendrogram zapisz do pliku postscriptowego.
2. Na danych `crabs` z biblioteki `MASS` porównaj metody klasyfikacji zaimplementowane w funkcjach `knn` i `nnet` z bibliotek `class` i `nnet`. Zakładając, że cecha przewidywana przyjmuje 4 wartości (*BF*, *BM*, *OF*, *OM*), wykonaj weryfikację krzyżową (5/6 danych do trenowania) i policz prawdopodobieństwo poprawnej predykcji dla obu metod.
- 3\*. Napisz funkcję rysującą wykres konturowy 2-wymiarowej gęstości  $f(x,y)$  tak, że warstwy ograniczają obszary o zadanych prawdopodobieństwach  $p=(p_1, \dots, p_k)$ . Na przykład, jeśli  $p = c(1,2,3,4)/5$ , to pierwsza warstwa jest zadana przez zbiór  $\{(x,y) : f(x,y) = z, P(f(x,y) \geq z) = 1/5\}$  dla pewnego  $z$ . Wykonaj rysunek na danych o zawartości pierwiastków w kilku rodzajach szkła:  

```
library(MASS); data(fg1); x=fg1[1:185,].
```

  
Policz gęstości (*Mg*, *Al*) za pomocą `kde2d` i narysuj warstwy dla danego wyżej  $p$ .