

Statystyka II Mat. Egzamin komputerowy 21.06.06

1. Wczytaj z archiwum *MacierzePot.zip* 12 potencjałów kontaktowych (PK) oznaczonych numerami nieparzystymi. PK są to symetryczne macierze 20 x 20 opisujące oddziaływania 210 par aminokwasów w białkach.
2. Napisz ogólną funkcję `upperVec`, która przekształca macierz na wektor złożony z elementów należących do jej górnej części (ewentualnie dolnej) oraz przekątnej. Ze zbioru wektorów $\mathbf{x}_1, \dots, \mathbf{x}_{12}$ otrzymanych z 12 PK, zbuduj macierz danych $X = [\mathbf{x}_1, \dots, \mathbf{x}_{12}]$. Czy w zbiorze X -- 210 obserwacji 12-wymiarowych są obserwacje odstające?
3. Wykonaj klasteryzację *average linkage*. Dendrogram zapisz do pliku *.ps*. Narysuj również oddziaływania między aminokwasami na płaszczyźnie dwóch pierwszych składowych głównych i policz udział ich wariancji w całkowitej wariancji danych. Ile jest klastrow (odp uzasadnij wykresem separowalności) ?
4. Porównaj metody klasyfikacji *lda*, *qda*, *multinom* (dyskryminacja logistyczna dla wielu klas z `library(nnet)`) oraz *nnet* w eksperymencie krosvalidacji sześciokrotnej na danych *iris*. Wykorzystaj wszystkie obserwacje. Cecha przewidywana przyjmuje 3 wartości. Jako miarę efektywności predykcji przyjmij wspólną informację.
5. Napisz odpowiednik, uogólnienie funkcji `lm` dla modelu $y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \varepsilon_i$ w którym $\varepsilon \sim N(0, \sigma^2 W)$, gdzie W jest znaną macierzą dodatnio określoną.
6. Zbuduj model regresji liniowej wielu zmiennych dla przefiltrowanych danych *Cars93*. Cechą przewidywaną jest `MPG.city`.

```
library(MASS)
X=na.omit(Cars93[,c(2,7,12:15,17,19,21,22,24,25)])
row.names(X)=X[,1]; X=X[,-1]
```