

Statystyka II Mat. Egzamin komputerowy

1. Wczytaj z archiwum *MacierzePot.zip* 23 potencjały kontaktowe (PK) zawierające energię oddziaływania między aminokwasami w białkach. (PK są to symetryczne macierze 20 x 20).
2. Napisz funkcję `upperVec`, która przekształca macierz na wektor złożony z elementów należących do jej górnej części. Policz macierz korelacji (CC) między takimi wektorami utworzonymi z macierzy PK. Czy w zbiorze wektorów x_1, \dots, x_{23} są obserwacje odstające?
3. Przekształć CC na macierz odległości euklidesowych i wykonaj klasteryzację *average linkage*. Dendrogram zapisz do pliku *.ps*. (PK powinny być na rysunku oznaczone numerami plików, z których zostały wczytane). Ile jest klastrów?
4. Porównaj metody klasyfikacji *nnet* i *qda* w eksperymencie krosvalidacji sześciokrotnej na danych *crabs* z biblioteki MASS. Cecha przewidywana przyjmuje 4 wartości: BF, BM, OF i OM. Policz prawdopodobieństwo poprawnej predykcji obu metod.
5. Plik *Orings.dat* zawiera dane NASA dotyczące awarii gumowych pierścieni uszczelniających silniki promów kosmicznych. Każdy prom miał 6 pierścieni. Pierwsza kolumna podaje temperaturę (w °F) w momencie startu, a druga – liczbę zepsutych pierścieni. Za pomocą modelu logistycznego oszacuj prawdopodobieństwo awarii co najmniej jednego pierścienia dla temperatury 37.5 °F.