

## Statystyka II Mat. Egzamin komputerowy 12.09.06

1. Wczytaj z archiwum *SMs.zip* 28 macierzy podstawień (MP) opisujących podobieństwo biochemiczne aminokwasów. MP są macierzami o wymiarze 20 x 20.
2. Przekształć macierze MP na wektory złożone z 400 elementów, następnie ze zbioru wektorów  $\mathbf{x}_1, \dots, \mathbf{x}_{28}$  zbuduj macierz danych  $X = [\mathbf{x}_1, \dots, \mathbf{x}_{28}]$ . Czy w zbiorze  $X$  są obserwacje odstające?
3. Wykonaj klasteryzację *k-średnich* danych  $X$ . Liczbę klastrów dobierz korzystając z wykresu separowalności. Narysuj dane  $X$  na płaszczyźnie dwóch pierwszych składowych głównych i przedstaw wynik klasteryzacji kolorując obserwacje.
4. Porównaj metody klasyfikacji *lda*, *qda*, *glm* (regresja logistyczna) w eksperymencie krosvalidacji sześciokrotnej na danych `MathAchSchool[, -c(1, 6)]` z biblioteki `nlme`. Cecha przewidywana  $y$  przyjmuje 2 wartości: „Catholic” i „Public”. Jako miarę efektywności predykcji przyjmij współczynnik Goodmana-Kruskala między  $y$  oraz  $y_{\text{pred}}$ .
5. Napisz funkcję `rnormMV(n, m, S)` generującą prostą próbę  $n$ -elementową z wielowymiarowego rozkładu normalnego o wektorze średnich  $\mathbf{m}$  i dodatnio określonej macierzy kowariancji  $\mathbf{S}$ . Dekompozycję  $\mathbf{S}$  wykonaj za pomocą rozkładu Choleskiego.
6. Zbuduj model analizy wariancji dla danych `quine` z biblioteki `MASS`. W modelu tym zmienną objaśnianą jest `Days` - liczba opuszczonych dni nauki przez australijskie dzieci. Dobierz transformację zmiennej `Days` korzystając z funkcji `boxcox` i `logtrans` oraz wykonaj automatyczną selekcję zmiennych.