

Statystyka II Mat. Egzamin komputerowy 2.09.08.

1. Wczytaj z archiwum `MacierzePot15_3kol.zip` potencjały kontaktowe (PK), czyli 12 symetrycznych macierzy 20×20 opisujących oddziaływania 210 par aminokwasów w białkach. PK są zapisane w plikach trójkolumnowych tak, że w pierwszej kolumnie znajduje się numer wiersza, w drugiej – numer kolumny, a w trzeciej – wartość elementu macierzy PK o tych współrzędnych. Napisz ogólną funkcję `three2Vec`, która przekształca plik trójkolumnowy na wektor złożony z elementów macierzy PK należących do jej dolnej części bez przekątnej. Ze zbioru wektorów $\mathbf{x}_1, \dots, \mathbf{x}_{12}$ otrzymanych z 12tu PK, zbuduj macierz danych $X = [\mathbf{x}_1, \dots, \mathbf{x}_{12}]$ (190 obserwacji 12-wymiarowych). Za pomocą funkcji `boxplot.stat` sprawdź, czy w zbiorze X są obserwacje odstające?
2. Porównaj metody klasyfikacji `lda`, `qda`, `glm` (regresja logistyczna) w eksperymencie krosvalidacji pięciokrotnej na danych `MathAchSchool[, -c(1, 6)]` z biblioteki `nlme`. Cecha przewidywana y przyjmuje dwie wartości: „Catholic” i „Public”. Jako miarę efektywności predykcji przyjmij wspólną informację między y oraz y_{pred} .
3. Zbuduj model regresji liniowej wielu zmiennych dla przefiltrowanych danych `Cars93`:
`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]`. Cechą przewidywaną jest `MPG.city`.
4. Dla modelu regresji z poprzedniego zadania:
 - policz (dowolną metodą) p-wartość testu F-Snedecora hipotezy: *współczynniki przy trzech ostatnich cechach są równe zero*;
 - policz p-wartość tej samej statystyki wykorzystując tylko $y = X[, 1]$, rozkład QR macierzy $X = \text{cbind}(1, X[, -1])$ oraz funkcję `pf`.
5. 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 $\{(x, y) : f(x, y) \leq z, P(f(x, y) \leq z) = 1/5\}$ dla pewnego z . Dla danych z zadania 3. policz gęstość (`MPG.city, EngineSize`) za pomocą `kde2d` i narysuj warstwy dla danego wyżej p .