

Case study.....pkt,
Pytania zamknięte.....pkt
Razem..... pkt

Imię i nazwisko.....

Ocena.....

K O L O K W I U M 3

31 maja 2019 r.

Statystyczna Eksploracja Danych

Pytania zamknięte (20 pytań po 0.5 pkt każde, jedna odpowiedź prawidłowa)

Zad 1. Czy optymalną liczbę skupień w metodzie k-średnich można wyznaczyć za pomocą wskaźnika \tilde{W} :

- (a) tak,
- (b) nie,
- (c) czasami.

Zad 2. Wybierz **prawdziwą** informację dotyczącą analizy czynnikowej (FA):

- (a) analiza czynnikowa jest tożsama z analizą składowych głównych,
- (b) wektor czynników ukrytych ma taki sam wymiar, co wektor obserwacji,
- (c) wektor czynników specyficznych ma taki sam wymiar, co wektor obserwacji.

Zad 3. Wybierz **prawdziwą** cechę skalowania wielowymiarowego:

- (a) nie może operować na macierzy odległości,
- (b) jest tożsamy z analizą składowych głównych, gdy macierz odmienności zastąpi się macierzą odległości,
- (c) nie jest stosowany do „obniżenia” pierwotnej wymiarowości układu

Zad 4. Algorytm k-średnich jest:

- (a) zawsze zbieżny i zawsze daje rozwiązanie globalne,
- (b) zawsze zbieżny ale nie zawsze daje rozwiązanie globalne,
- (c) czasami zbieżny.

Zad 5. Zaznacz nieprawdziwą informację dotyczącą analizy składowych głównych:

- (a) wektor i-tej składowej głównej jest równy i-temu wektorowi własnemu macierzy kowariancji S ,
- (b) zamiast macierzy kowariancji można użyć macierzy korelacji,
- (c) kolejne składowe główne są ze sobą skorelowane.

Case study (5 pkt)

Dla poniższych danych przeprowadź analizę składowych głównych:

- wyznacz składowe główne,
- wyznacz wariancje rzutów obserwacji na poszczególne składowe główne,
- określ wartości współrzędnych w nowym układzie (wyznaczonym przez składowe główne),
- oceń, które składowe można pominąć i dlaczego.

$$X = \begin{pmatrix} 1 & 9 & 8 \\ 2 & 8 & 8 \\ 3 & 7 & 8 \\ 4 & 6 & 8 \\ 5 & 5 & 8 \end{pmatrix}$$