

Przetwarzanie i analiza dużych zbiorów danych - zadanie 3

Paweł Ciupka 234048

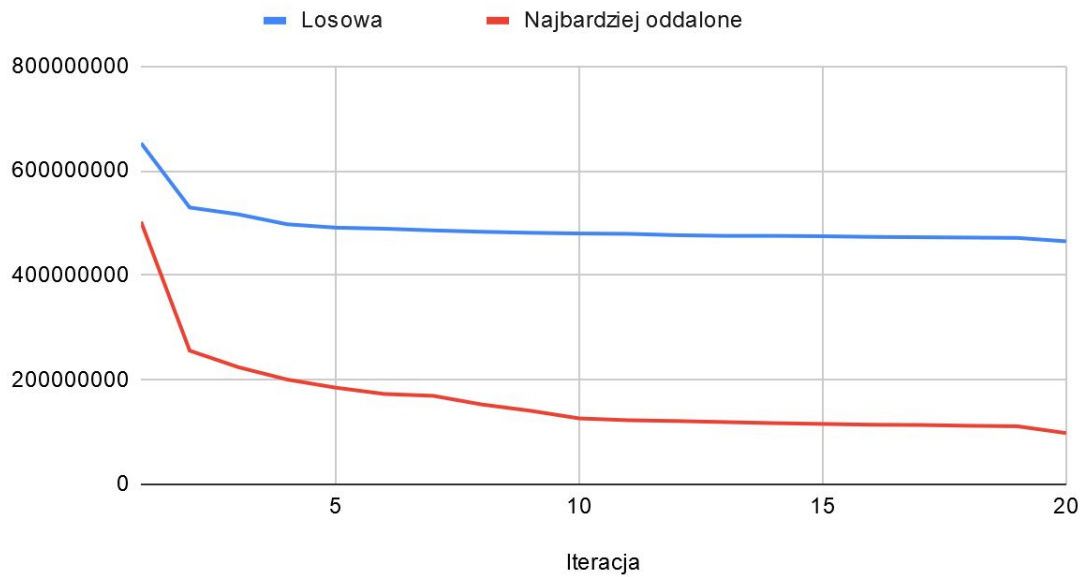
Bartosz Łuniewski 234086

Filip Woźniak 234131

Wykresy funkcji kosztu

Miara euklidesowa

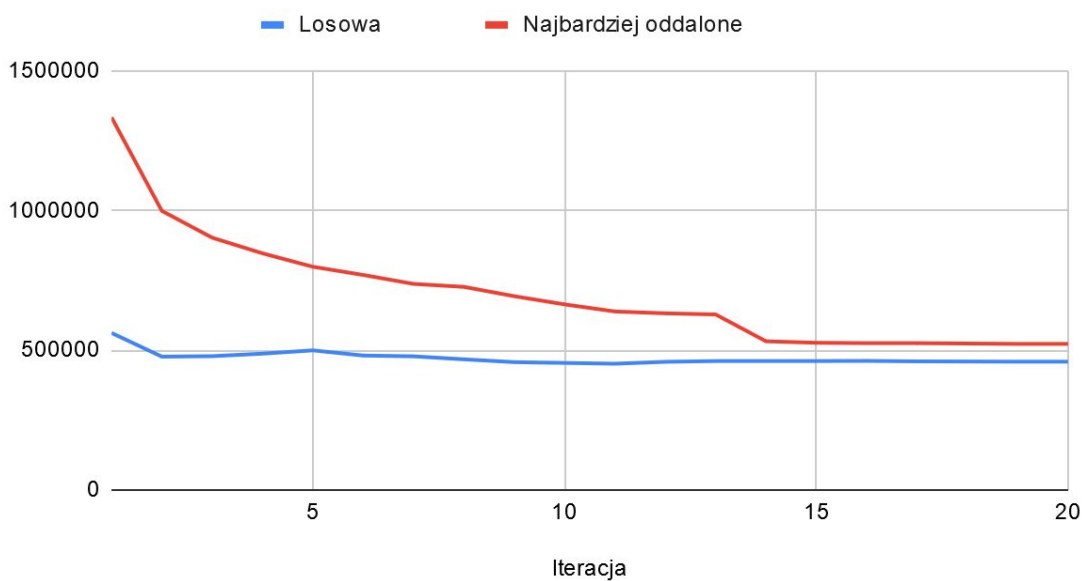
Wykres dla miary euklidesowej



Rys. 1. Wykres dla miary euklidesowej

Miara Manhattan

Miara Manhattan



Rys. 2. Wykres dla miary Manhattan

Tabele danych zawartych na wykresach funkcji kosztu

Tab. 1. Koszt w poszczególnych iteracjach

Iteracja	Miara Euklidesowa		Miara Manhattan	
	Losowa (3b.txt)	Najbardziej oddalone (3c.txt)	Losowa (3b.txt)	Najbardziej oddalone (3c.txt)
1	653102132,1	502572710,7	561843,1	1333165,5
2	529665829,1	255809509,7	477004,7919	998540,9595
3	516603786,5	224147429	478690,3212	902553,0435
4	497663824,7	200526511,6	487679,8108	846560,0925
5	491132803,4	184946757,5	499478,5159	798566,1345
6	489173497	172885012,5	480937,6936	769236,4935
7	485907986,3	169367003,5	478128,4781	737240,5215
8	483295577,8	152782104	467453,4592	726575,1975
9	481336271,4	140720359	457340,2834	693246,06
10	480030067,1	126145750,4	454531,0679	663916,419
11	479376965	122627741,4	451721,8524	638586,2745
12	476764556,4	121120023,3	458463,9696	631920,447
13	475458352,2	119109732,4	461273,1851	627920,9505
14	475458352,2	117099441,6	461273,1851	531933,0345
15	474805250,1	115591723,5	461273,1851	526600,3725
16	473499045,8	114084005,3	461835,0282	525267,207
17	472845943,7	113581432,6	460149,4989	525267,207
18	472192841,5	112073714,5	459587,6558	523934,0415
19	471539739,4	111068569,1	459025,8127	522600,876
20	465008718,1	98001678,58	459025,8127	522600,876

Tab 2. Procentowe zmiany kosztu w poszczególnych iteracjach

Iteracja	Miara Euklidesowa		Miara Manhattan	
	Losowa (3b.txt)	Najbardziej oddalone (3c.txt)	Losowa (3b.txt)	Najbardziej oddalone (3c.txt)
1	0	0	0	0
2	18,9	49,1	15,1	25,1
3	20,9	55,4	14,8	32,3
4	23,8	60,1	13,2	36,5
5	24,8	63,2	11,1	40,1
6	25,1	65,6	14,4	42,3
7	25,6	66,3	14,9	44,7
8	26	69,6	16,8	45,5
9	26,3	72	18,6	48
10	26,5	74,9	19,1	50,2
11	26,6	75,6	19,6	52,1
12	27	75,9	18,4	52,6
13	27,2	76,3	17,9	52,9
14	27,2	76,7	17,9	60,1
15	27,3	77	17,9	60,5
16	27,5	77,3	17,8	60,6
17	27,6	77,4	18,1	60,6
18	27,7	77,7	18,2	60,7
19	27,8	77,9	18,3	60,8
20	28,8	80,5	18,3	60,8

Wnioski

W przypadku miary Euklidesowej lepszym rozwiązaniem było wykorzystanie z maksymalnie oddalonych od siebie centroidów. Określenie maksymalnie oddalonych od siebie punktów, skutkuje dokładniejszym wynikiem przeprowadzonego badania. Wyniki otrzymane przy pomocy losowo wybranych centroidów, są mocno uzależnione od początkowego rozłożenia punktów. Podczas badania miary Manhattan, lepszym rozwiązaniem było wykorzystanie losowo wybranych centroidów. Zauważono wypłaszczenie wyników zastosowanych metryk po 10 iteracji w każdym z przeprowadzonych badań.