

Algorytmy Numeryczne – Zadanie 3

Metoda ALS w systemach Rekomendacji

Kasper Cisewski 253902

Marcel Dajnowicz 253971

1. Wstęp

Zadaniem było zaimplementowanie metody ALS korzystając z gotowej metody Gaussa z drugiego zadania.

2. Parsowanie Danych

Do testów użyliśmy prawdziwych danych o produktach i rekomendacjach ze sklepu Amazon. Plik zawiera rekordy o 500 000 produktach i ponad 7 mln ocen. Po wstępnym parsowaniu otrzymaliśmy dane takie, jak id produktu czy klienta oraz ocena i kategoria przedmiotu. Na tak przygotowanych danych, rozpoczynamy proces wyboru odpowiednich klientów i ich recenzji. Wybieramy produkty, które otrzymały najwięcej ocen, a z nich użytkowników, którzy ocenili podane produkty najwięcej razy. Z tak otrzymanych danych otrzymujemy macierz składająca się z ocen dla podanych produktów przez konkretnych użytkowników.

```
[0, 4, 4, 4, 4, 4, 0, 0, 4, 4],
[0, 5, 5, 5, 5, 5, 0, 0, 5, 5],
[4, 0, 0, 0, 0, 0, 4, 4, 0, 0],
[4, 0, 0, 0, 0, 0, 4, 4, 0, 0],
[0, 5, 5, 5, 5, 5, 0, 0, 5, 5],
[5, 0, 0, 0, 0, 0, 5, 5, 0, 0],
[4, 0, 0, 0, 0, 0, 4, 4, 0, 0],
[0, 5, 5, 5, 5, 5, 0, 0, 5, 5],
[5, 0, 0, 0, 0, 0, 5, 5, 0, 0],
[4, 0, 0, 0, 0, 0, 4, 4, 0, 0]
```

3. Generowanie i Opracowanie Wyników

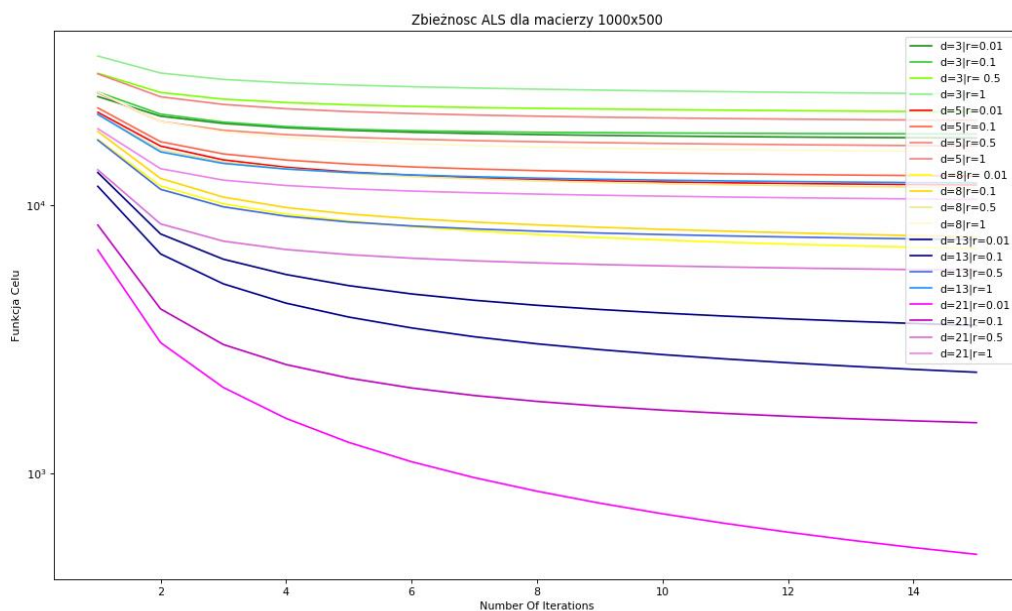
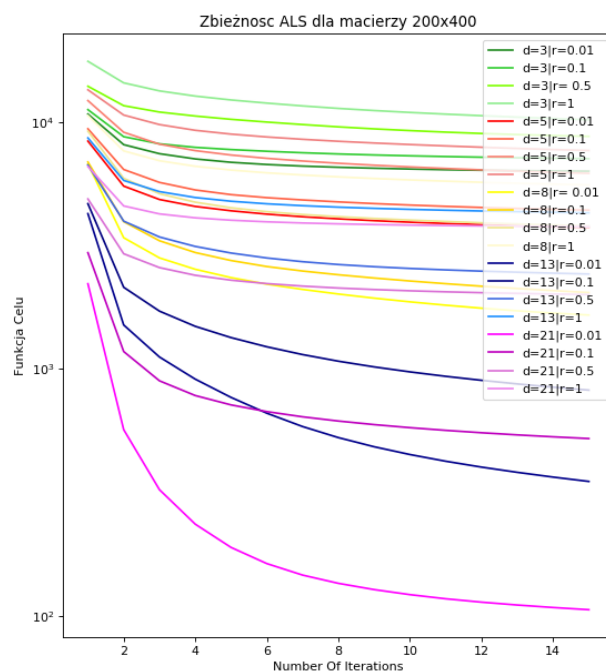
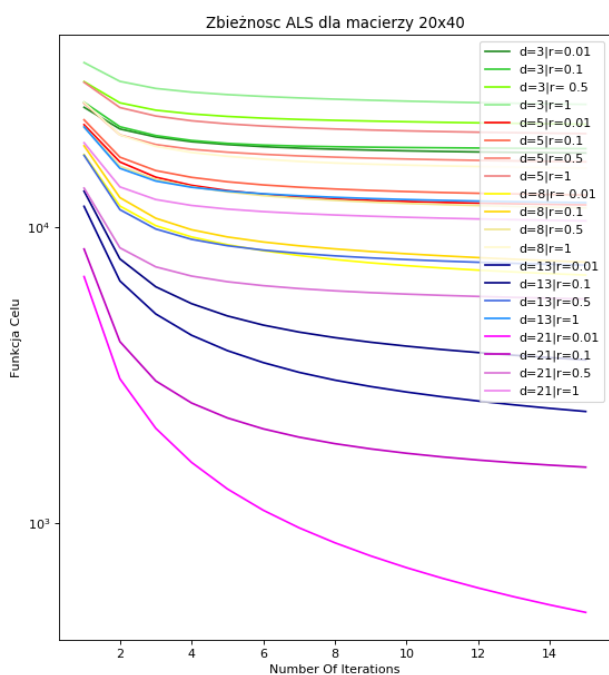
Po napisaniu implementacji ALS rozpoczęliśmy testy dla macierzy składającej się z 10 użytkowników i 10 produktów. Macierz została podzielona na zbiór uczący i zbiór testowy.

0.00 3.30	0.00 2.08	2.00 2.00	5.00 4.90	0.00 3.30	0.00 3.30	2.00 1.99	0.00 2.78	0.00 4.82	0.00 0.60
0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00	1.00 0.00
0.00 3.02	4.00 3.91	1.00 1.02	0.00 1.95	0.00 3.02	0.00 3.02	1.00 1.00	5.00 2.36	2.00 2.00	0.00 0.58
0.00 1.74	0.00 1.63	5.00 4.81	0.00 1.97	0.00 1.74	0.00 1.74	5.00 4.90	0.00 2.77	0.00 1.92	0.00 0.12
0.00 0.73	0.00 0.68	2.00 1.96	0.00 0.82	0.00 0.73	0.00 0.73	2.00 2.00	0.00 1.14	0.00 0.80	0.00 0.05
5.00 4.97	5.00 5.03	5.00 4.96	5.00 5.04	5.00 4.97	5.00 4.97	5.00 5.01	5.00 4.95	5.00 5.00	1.00 0.79
1.00 0.99	1.00 1.00	0.00 0.52	0.00 0.99	1.00 0.99	1.00 0.99	0.00 0.52	0.00 0.82	0.00 0.99	0.00 0.18
0.00 3.31	2.00 2.02	2.00 2.00	5.00 5.00	0.00 3.31	0.00 3.31	2.00 1.99	0.00 2.78	5.00 4.91	0.00 0.60
0.00 4.89	5.00 4.95	5.00 4.95	5.00 4.96	0.00 4.89	0.00 4.89	5.00 5.00	3.00 4.89	0.00 4.92	0.00 0.77
5.00 4.96	5.00 5.00	1.00 1.05	5.00 5.00	5.00 4.96	5.00 4.96	1.00 1.00	5.00 3.60	5.00 5.00	1.00 0.99

Po lewej strony każdej komórki została zaznaczona wartość odpowiadająca ocenie danego produktu, a po prawej wynik predykcji. Komórki na różowo nie zostały przekazane do zbioru uczącego co dało nam możliwość sprawdzenia algorytmu na danych, które wcześniej nie widział. Jak można zauważyć, wyniki są zbliżone do podanych co potwierdza prawidłowa implementacja.

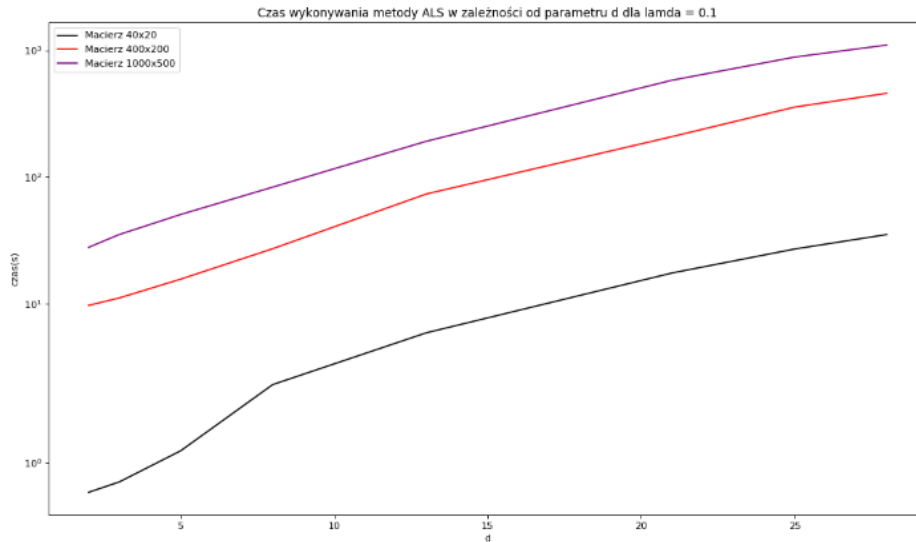
4. Tempo zbieżności metody ALS w zależności od obranych parametrów

Na podstawie wykresów badających tempo zbieżności, zauważyć możemy, że większe D oraz mniejsza λ dają mniejsze, tym samym lepsze wyniki. Zauważyć można też bardzo duży wpływ ilości iteracji dla funkcji celu. Im jest ich więcej tym wyniki są dokładniejsze. Rozmiar danych dla małego zbioru danych to 20 produktów i 40 użytkowników, 200 i 400, 1000 i 500 kolejno dla średniego i dużego zbioru.

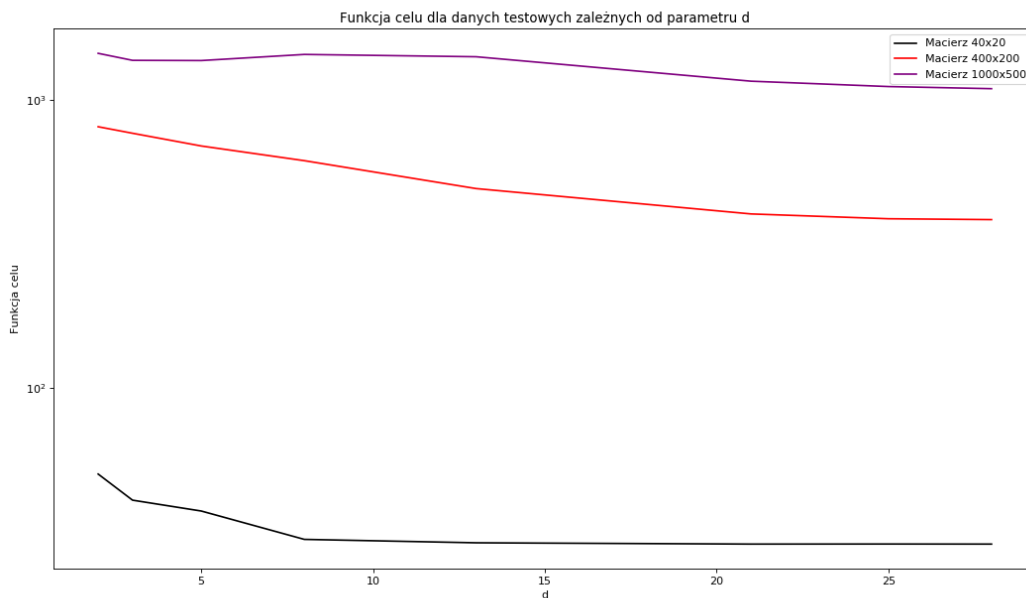


5. Wpływ parametru D na jakość stworzonych rekomendacji i czas obliczeń

Poniższe testy zostały wykonane podczas 100 iteracji i λ ustawionej na 0.1. Jak możemy zauważyć, im większe D , tym proces wymaga więcej czasu przez dodatkowe obliczenia.



Poniższy wykres pokazuje brak większego wpływu D na wyniki funkcji celu od pewnego momentu. Zalecane jest znalezienie takiego D , które od pewnego momentu nie daje większych różnic w wynikach i tym samym dłuższego czasu obliczeń.



Przy właściwym dobraniu D , λ , odpowiednią ilością iteracji i wraz z dobrze wypełnioną bazą danych ocenami użytkowników, otrzymać możemy dobre predykcje produktów, które nie wymagają dużo czasu.