

Trabalho 2

Aprendizado de Máquina

Sistema de Recomendação de Filmes

Bruno Lippert e Raul Rodrigues

MovieLens

- GroupLens
- MovieLens dataset
- Small educational dataset
- 100,000 ratings
- 9,000 movies
- **600** users

Solução

1. MovieLens GroupLens dataset - small educational
2. Processamento inicial do dataset
3. Modelo ou sistema que recebe um novo usuário
4. Recomendações de filmes são criadas

Processamento inicial

ratings.csv

userId,movieId,rating,timestamp

1,1,4.0,964982703

1,3,4.0,964981247

1,6,4.0,964982224

1,47,5.0,964983815

1,50,5.0,964982931

1,70,3.0,964982400

1,101,5.0,964980868

1,110,4.0,964982176

1,151,5.0,964984041

1,157,5.0,964984100

1,163,5.0,964983650

1,216,5.0,964981208

1,223,3.0,964980985

.

.

.

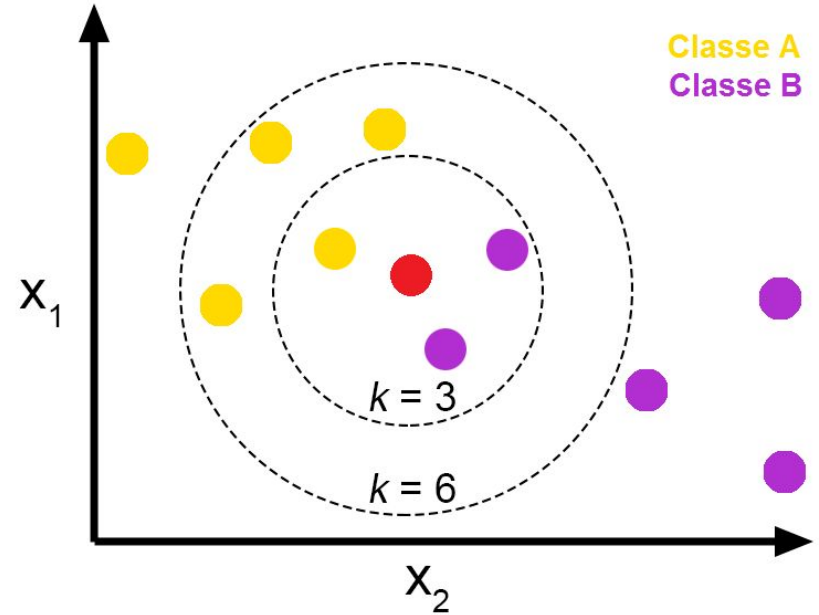
Processamento inicial

610 rows × 9742 columns

userId	1	2	3	...	193583	193585	193587
0	10	0	10	...	0	0	0
1	0	0	0	...	0	0	0
2	0	0	0	...	0	0	0
3	0	0	0	...	0	0	0
4	10	0	0	...	0	0	0
...
605	-5	0	0	...	0	0	0
606	10	0	0	...	0	0	0
607	-5	-10	-10	...	0	0	0
608	0	0	0	...	0	0	0
609	20	0	0	...	0	0	0

Modelo

- scikit-learn nearest neighbors



Demo

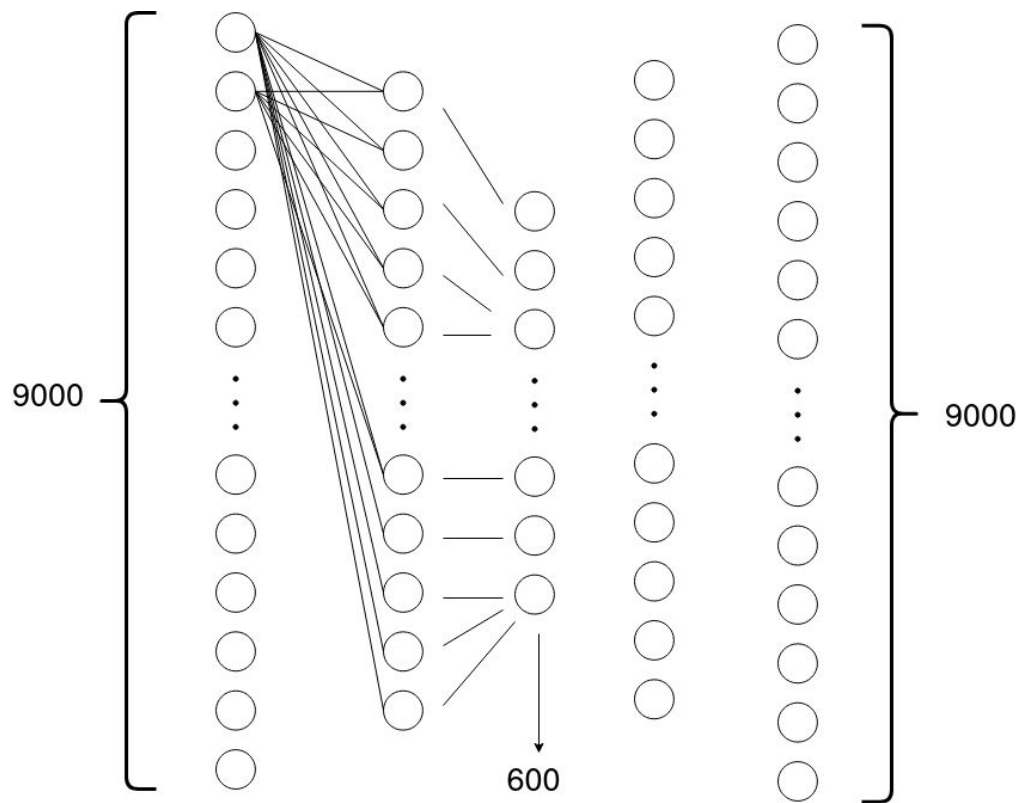
Problemas

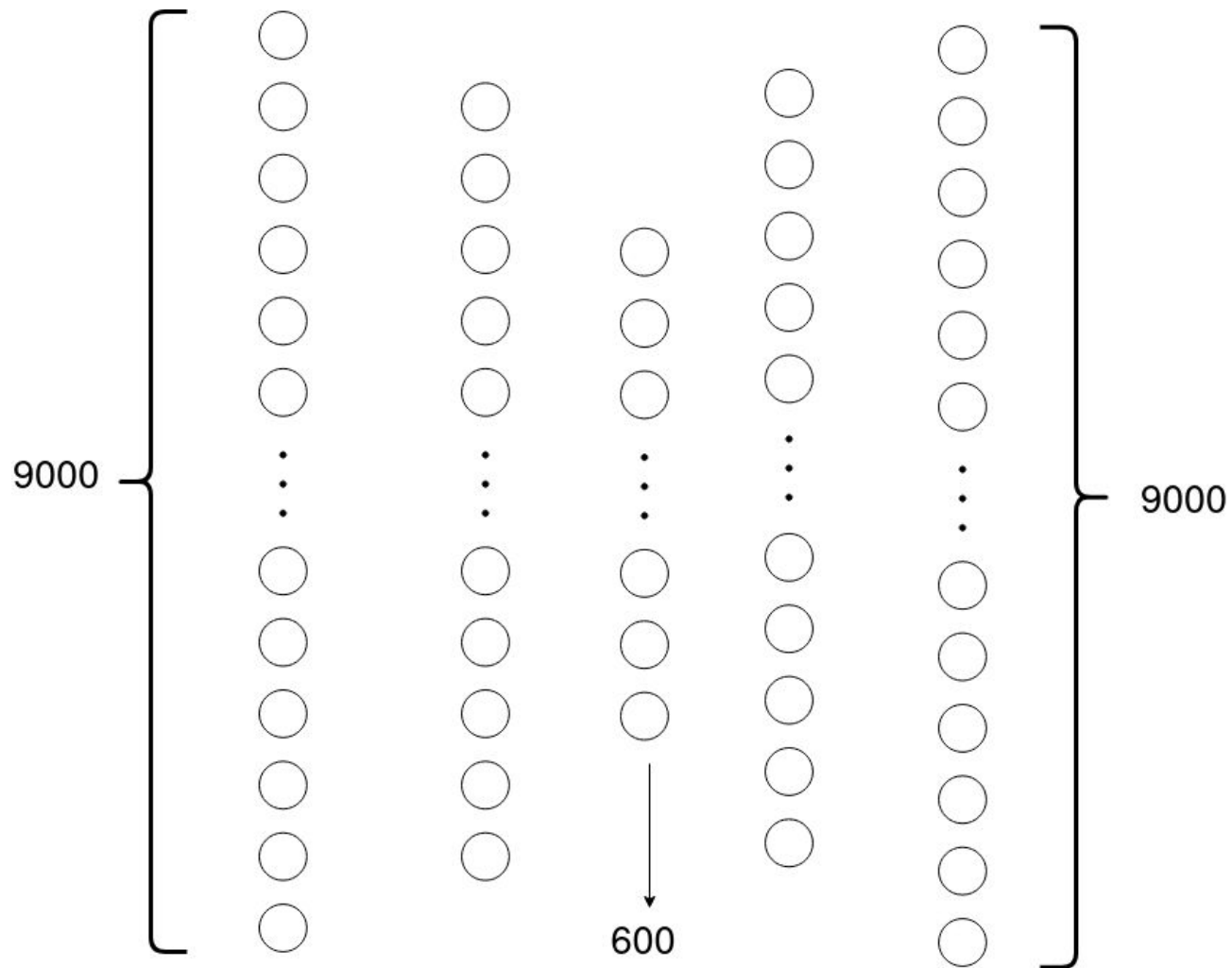
- Pouca variação com a entrada
- Tamanho do dataset
- Problemas fundamentais de um sistema feito em cima de um dataset

Outras tentativas

- Autoencoder
- Matriz de correlação

Autoencoder





Matriz de correlação

610 rows × 9742 columns

	1	2	3	4	5	...	193567	193609
1	1	0.13101	0.121416	0.004797		...	-0.019872	-0.019872
2	0.13101	1	0.133913	-0.079404		...	-0.007734	-0.007734
3	0.121416	0.133913	1	0.004788		...	-0.002861	-0.002861
4	0.004797	-0.079404	0.004788	1		...	0.002748	0.002748
5	0.047205	0.129528	0.244293	0.045817		...	-0.000911	-0.000911
...
193581	-0.019872	-0.007734	-0.002861	0.002748		...	1	-0.002861
193583	-0.019872	-0.007734	-0.002861	0.002748		...	0.004788	1
193585	-0.019872	-0.007734	-0.002861	0.002748		...	1	1
193587	-0.019872	-0.007734	-0.002861	0.002748		...	1	1
193609	-0.019872	-0.007734	-0.002861	0.002748		...	-0.001642	1

Trabalho futuro

- Usar dataset maiores
- Usar gênero dos filmes
 - Cálculo de distância
 - Score em nível de contagem de gêneros
- Criação de arquétipos

Trabalho 2

Aprendizado de Máquina

Sistema de Recomendação de Filmes

Bruno Lippert e Raul Rodrigues