1 Dataset

In this project, we only need to consider the ratings dataset. And this dataset has four columns, userId, movieId, rating, timestamp. The first five records are:

	userId	movieId	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931

Construct the rating matrix which is $n \times p$. And in this dataset, $n = 610, p = 9724$. In the following of dataset split, we divide original dataset into two parts 90% of training data and 10% of testing data.

2 Learning

objective function:

$$\min_{U,V} F(U,V) := \frac{1}{2} \sum_{(i,j)\in\Omega_1} (M_{i,j} - u_i v_j^T)^2 + \frac{\lambda}{2}(||U||_F^2 + ||V||_F^2)$$

Q1. We assume that $U \in R^{n\times k}, V \in R^{m\times k}$ and k is the number of latent factors. Then,

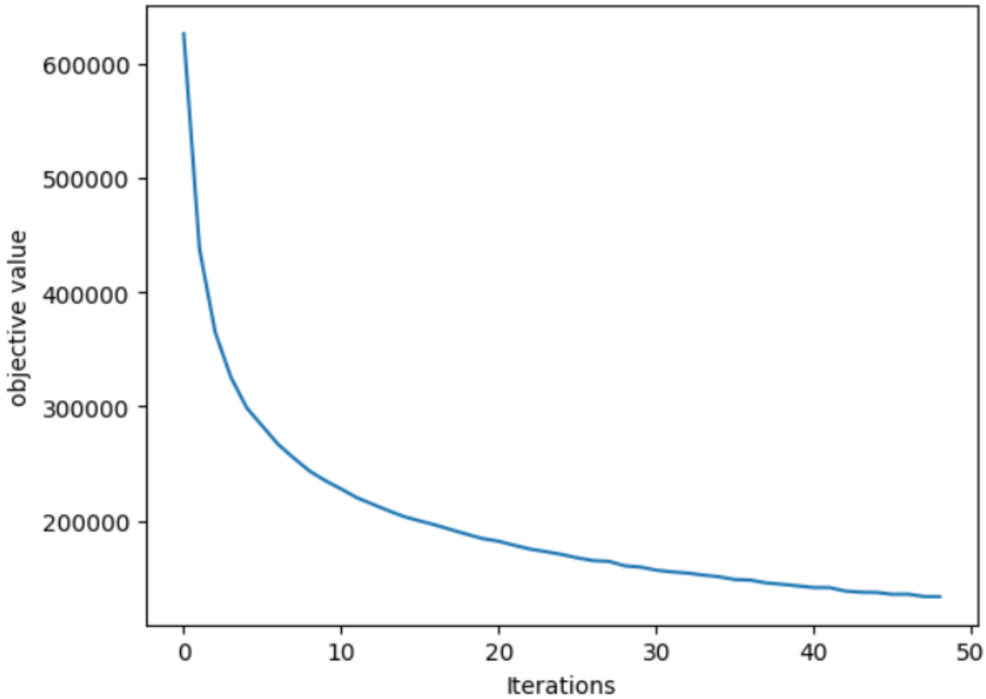
$$\frac{\partial F}{\partial u_i} = \sum_{j\in[m]:(i,j)\in\Omega_1} (u_i v_j^T - M_{ij})v_j + \lambda u_i = u_i(\sum_{j\in[m]:(i,j)\in\Omega_1} v_j^T v_j + \lambda I) - \sum_{j\in[m]:(i,j)\in\Omega_1} M_{ij} v_j$$
$$\frac{\partial F}{\partial v_j} = \sum_{i\in[n]:(i,j)\in\Omega_1} (u_i v_j^T - M_{ij})u_i + \lambda v_j = v_j(\sum_{i\in[n]:(i,j)\in\Omega_1} u_i^T u_i + \lambda I) - \sum_{i\in[n]:(i,j)\in\Omega_1} M_{ij} u_i$$

Q2. For a given index ***(i,j)*** and $\lambda = 1$,SGD update rules are:

$$u_i = u_i - \eta[(u_i v_j^T - M_{ij})v_j + u_i]$$
$$v_j = v_j - \eta[(u_i v_j^T - M_{ij})u_i + v_j]$$

where η is the learning rate.

Q3. In this step, we set the total iteration equal to 50. And the plotting of objective value against the number of iterations is shown below:



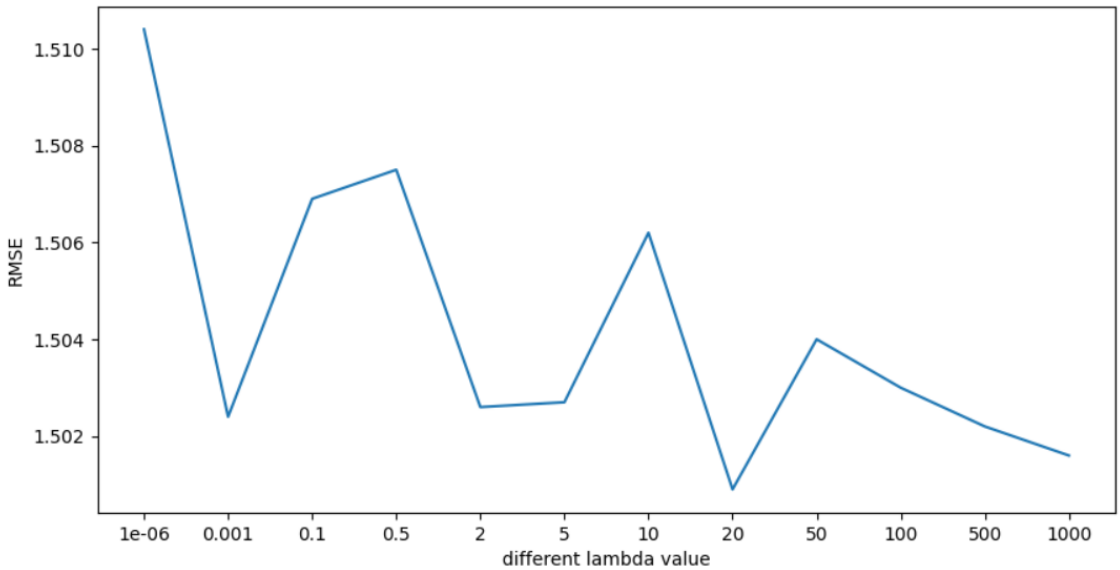
From the figure, we can find that as the iteration increase the objective function value decreases. And the value decreases rapidly In first 10 iterations and then decreases slowly. It shows that SGD can quickly find a local minimum of $F(U, V)$.

3 Evaluation

Q1. when $\lambda = 1$, RMSE value is 1.5189.

Q2. pick λ from $\{10^{-6}, 10^{-3}, 0.1, 0.5, 2, 5, 10, 20, 50, 100, 500, 1000\}$. The RMSE values are:

```
[1.5104, 1.5024, 1.5069, 1.5075, 1.5026, 1.5027, 1.5062, 1.5009, 1.504, 1.503,
1.5022, 1.5016]
```



From the plotting above, we can find too small or big λ value will cause large RMSE. Therefore, it need to choose a proper for λ . And we can find that when $\lambda = 20$, the corresponding RMSE achieves minimal RMSE.