Recommender systems
SD-TSIA 211

Teachers : Anas Barakat, Olivier Fercoq, Nidham Gazagnadou, Ekhine Irurozki, Kimia Nadjahi
18 December 2020

You can do the computer lab alone or in pairs. Please write a report and post it on `e-campus`. You can do it as a jupyter notebook or a pdf file.

Then, each of you will have to evaluate a couple of other students' reports and give comments.

Only the fact that you produce a report and evaluate your peers count in the final grade, so do not worry if you do not finish everything.

1 Presentation of the model

U is the set of users, I is the set of items (here, films). For each couple (u, i), either user u has not watched film i and we do not have any datum, or we have a grade $R_{u,i}$ for film i by user u.

The model presented in [KBV09] makes the assumption that there exists a joint latent feature space F such that user-item interactions are inner products in that space. According to this model, one should have $R_{u,i} \approx \sum_{f \in F} Q_{u,f} P_{f,i}$ where $Q_{u,:}$ is a representation of user u in the feature space and $P_{:,i}$ is a representation of item i in the feature space. The strength of this model is to predict a probable grade that user u would give film i if she ever watched it. Thus, we can recommend her films that she has not watched but that she may enjoy.

We then train the model using regularized least squares :

$$\begin{aligned}(\hat{P}, \hat{Q}) &= \arg\min_{P,Q} \frac{1}{2} \sum_{(u,i) \in K} \left(R_{u,i} - \sum_{f \in F} Q_{u,f} P_{f,i} \right)^2 + \frac{\rho}{2} \left(\sum_{u \in U, f \in F} Q_{u,f}^2 + \sum_{i \in I, f \in F} P_{f,i}^2 \right) \qquad (1)\\ &= \arg\min_{P,Q} \ \frac{1}{2} \|1_K \circ (R - QP)\|_F^2 + \frac{\rho}{2} \|Q\|_F^2 + \frac{\rho}{2} \|P\|_F^2\end{aligned}$$

where K is the set of couples (u, i) for which $R_{u,i}$ is known, $\|\cdot\|_F$ is Frobenius's norm, $(1_K)_{u,i} = 0$ if $(u, i) \notin K$ and $(1_K)_{u,i} = 1$ if $(u, i) \in K$, $(A \circ B)_{u,i} = A_{u,i} B_{u,i}$ and $\rho > 0$ is a regularization parameter. We shall take here $\rho = 0.3$ and $F = \{0, 1, 2, 3\}$.

When $\rho = 0$ and $K = U \times I$, the solution of this problem is a truncated singular value decomposition. In the general case, we need to use an optimization algorithm.

Question 1.1
Download the Movielens database [HKBR99] on the website :

http://files.grouplens.org/datasets/movielens/ml-100k.zip.

Run the function `load_movielens` of `movielens_utils.py` with the correct file name and check that the matrix R has size $943 \times 1\,682$. What is the `minidata` option doing ?

Question 1.2
How many user and films are there in the database ? What is the total number of grades ?

2 Find P when Q^0 is fixed

We initialize the algorithm with Q^0 (resp. P^0) defined as the left (resp. right) singular vectors of R associated to the $|F|$ largest singular values. The missing values of R will be fixed to 0 for this initialization phase. You may use the function `scipy.sparse.linalg.svds`.

First, we want to solve the following simpler problem :

$$
\begin{aligned}
g(P) &= \frac{1}{2}\|1_K \circ (R - Q^0 P)\|_F^2 + \frac{\rho}{2}\|Q^0\|_F^2 + \frac{\rho}{2}\|P\|_F^2 \\
P^1 &= \arg\min_P \; g(P)
\end{aligned}
$$

Question 2.1
Calculate the gradient of function g. We will admit that this gradient is Lipschitz continuous with constant $L_0 = \rho + \|(Q^0)^\top Q^0\|_F$.

Question 2.2
The function `objective` provided in the file `movielens_utils.py` computes $g(P)$. Complete this function so that it also computes $\nabla g(P)$. You may check your calculations with the function `scipy.optimize.check_grad` (you may need `numpy.reshape` and `numpy.ravel` because `check_grad` does not accept matrix variables).

Question 2.3
Code a function `gradient(g, P0, gamma, epsilon)` that minimizes a function g using the gradient method with a constant step size γ, starting from the initial point P^0 and with stopping criterion $\|\nabla g(P_k)\|_F \leq \epsilon$.

Question 2.4
Run the function coded in the previous question in order to minimize g up to the precision $\epsilon = 1$.

Question 2.5
Add a line search to your gradient method, so that you do not rely on the Lipschitz constant of the gradient any more.

3 Resolution of the full problem

Question 3.1
Let f be the function defined by $f(P,Q) = \frac{1}{2}\|1_K \circ (R - QP)\|_F^2 + \frac{\varrho}{2}\|Q\|_F^2 + \frac{\varrho}{2}\|P\|_F^2$.

By remarking that f is a polynomial of degree 4, show that its gradient is not Lipschitz continuous.

Question 3.2
Solve Problem (1) by the gradient method with line search until reaching the precision $\|\nabla f(P_k, Q_k)\|_F \leq \epsilon$ with $\epsilon = 100$. How do you interpret what the algorithm returns ?

Question 3.3
What film would you recommend to user 300 ?

Références

[HKBR99] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *Proc. of 22nd ACM SIGIR conference*, pages 230–237. ACM, 1999.

[KBV09] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8) :30–37, 2009.