Matrix factorization optimization

October 24, 2016

The matrix factorization model is

$$R \approx PQ^\mathsf{T},$$

where R is the $m \times n$ user-item rating matrix, P is the $m \times k$ user factor matrix and Q is the $n \times k$ item factor matrix (k is the number of latent vairables/the dimension of the latent space).

Let

$$f(P,Q) = \sum_{(u,i)\in S} (r_{ui} - p_u q_i^\mathsf{T})^2 + \lambda(\|P\|_F^2 + \|Q\|_F^2) \tag{1}$$

be the objective function, that is, we are going to minimize f so as to solve for P and Q:

$$(\hat{P},\hat{Q}) = \arg\min_{P,Q} f(P,Q), \tag{2}$$

where p_u is the $1\times k$ user factor for user u (the u-th row in P), q_i is the $1\times k$ item factor for item i (the i-th row in Q), $p_u q_i^\mathsf{T} = \sum_{h=1}^{k} p_u(h) \times q_i(h)$ (i.e., the dot product of row vector p_u and q_i), $\|P\|_F^2 = \sum_{j=1}^{m}\sum_{h=1}^{k} P(j,h)^2$, $\|Q\|_F^2 = \sum_{j=1}^{n}\sum_{h=1}^{k} Q(j,h)^2$, and S is the set of ratings (i.e., $\sum_{(u,i)\in S}$ is sum over all the user-item ratings. In other words, $\sum_{(u,i)\in S}$ is sum over all u and i where $R(u,i) \neq 0$).

1 Closed-form solution

To solve the optimization problem in Equation 2, we use alternating least squares (ALS):

- Fix Q and solve for P. The objective function becomes

$$\begin{aligned}\min_P f_P(P) = & \sum_{(u,i)\in S} (r_{ui} - p_u q_i^\mathsf{T})^2 + \lambda\|P\|_F^2 \\ = & \sum_{(u,i)\in S} (r_{ui} - p_u q_i^\mathsf{T})^2 + \lambda \sum_u \|p_u\|_2^2\end{aligned} \tag{3}$$

where $\|p_u\|_2^2 = \sum_{h=1}^{k} p_u(h)^2$.

Take the derivative of $f_P(P)$ with respect to p_u and set it to 0,

$$\frac{\partial f_P(P)}{\partial p_u} = -2 \sum_{i|(u,i)\in S} (r_{ui} - p_u q_i^{\mathsf{T}}) q_i + 2\lambda p_u = 0 \tag{4}$$

where $\sum_{i|(u,i)\in S}$ means sum over all the items that user u has ever rated.

Then we get

$$\begin{aligned} -\sum_{i|(u,i)\in S} (r_{ui} - p_u q_i^{\mathsf{T}}) q_i + \lambda p_u = 0 \\ \sum_{i|(u,i)\in S} q_i r_{ui} = p_u(\lambda \mathbf{I} + \sum_{i|(u,i)\in S} q_i^{\mathsf{T}} q_i) \end{aligned} \tag{5}$$

and thus

$$p_u = (\sum_{i|(u,i)\in S} q_i r_{ui})(\lambda \mathbf{I} + \sum_{i|(u,i)\in S} (q_i^{\mathsf{T}} q_i))^{-1} \tag{6}$$

Equation 6 (actually [p_u]) gives the closed-form solution to the problem in Equation 3

- Fix P and solve for Q. It is almost the same procedure as for solving for P.

$$q_i = (\sum_{u|(u,i)\in S} p_u r_{ui})(\lambda \mathbf{I} + \sum_{u|(u,i)\in S} (p_u^{\mathsf{T}} p_u))^{-1} \tag{7}$$

- Iterate the above two steps until convergence (i.e., no significant change in P and Q or f)

2 Gradient-descent solution

- for each $(u,i) \in S$

  Take the derivative of $f_P(P)$ with respect to p_u,

$$\frac{\partial f_P(P)}{\partial p_u} = -2 \sum_{i|(u,i)\in S} (r_{ui} - p_u q_i^{\mathsf{T}}) q_i + 2\lambda p_u \tag{8}$$

p_u will be updated as follows:

$$\begin{aligned} p_u &= p_u - \frac{\gamma}{2}(-2 \sum_{i|(u,i)\in S} (r_{ui} - p_u q_i^\intercal) q_i + 2\lambda p_u) \\ &= p_u + \gamma \sum_{i|(u,i)\in S} (r_{ui} - p_u q_i^\intercal) q_i - \gamma\lambda p_u \\ &= (1-\gamma\lambda) p_u + \gamma \sum_{i|(u,i)\in S} (r_{ui} - p_u q_i^\intercal) q_i \end{aligned} \tag{9}$$

where $\frac{\gamma}{2}$ is the learning rate.
Similarly,

$$\begin{aligned} q_i &= q_i - \frac{\gamma}{2}(-2 \sum_{u|(u,i)\in S} (r_{ui} - p_u q_i^\intercal) p_u + 2\lambda q_i) \\ &= q_i + \gamma \sum_{u|(u,i)\in S} (r_{ui} - p_u q_i^\intercal) p_u - \gamma\lambda q_i \\ &= (1-\gamma\lambda) q_i + \gamma \sum_{u|(u,i)\in S} (r_{ui} - p_u q_i^\intercal) p_u \end{aligned} \tag{10}$$