

## PS 3

### Question 2: Bi-Linear Models and SGD

Your Name: Gino Rospigliosi

People I discussed this question with: None

## Bilinear Models for Recommendation ¶

For this problem, you will derive a very simple recommendation system that uses a combination of unsupervised and supervised approaches and demonstrates use of Stochastic Gradient Descent.

Specifically, in class we discussed recommender models of the form:

$$f(user, movie) = \langle v_u, v_m \rangle + b_u + b_m + \mu$$

where  $v$  is a vector that represents a user's or movie's location in an N-Dimensional space,  $b$  is a vector that represents a specific "bias" term for each movie and user, and  $\mu$  is a scalar that represents a kind of global anchor or base score (i.e., a sort of average movie rating). This means that each user has two vectors (e.g.,  $v_{\text{jack smith}}$  and  $b_{\text{jack smith}}$ ), and each movie has two vectors (e.g.,  $v_{\text{Avengers}}$  and  $b_{\text{Avengers}}$ ), with each of those vectors being N-Dimensional (in class we used two dimensions). For this, we constructed a loss function as follows:

$$Cost = Loss + Penalty$$

where

$$Loss = \sum_{(u,m) \in \text{Ratings}} \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m})^2$$

and

$$Penalty = \frac{\lambda}{2} \left[ \sum_u ||v_u||^2 + b_u^2 + \sum_m ||v_m||^2 + b_m^2 \right]$$

## Task 1: Analytical Gradients

To use stochastic gradient descent, we first need to write down the gradients. Using the above cost function (including both the loss and penalty), compute the following partial derivatives:

$$\begin{aligned}\frac{\partial \text{Cost}}{\partial v_u} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial v_u} + \frac{\lambda}{2} [\sum_u 2 * ||v_u||] * \frac{\partial}{\partial v_u} \\ \frac{\partial \text{Cost}}{\partial v_m} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial v_m} + \frac{\lambda}{2} [\sum_m 2 * ||v_m||] * \frac{\partial}{\partial v_m} \\ \frac{\partial \text{Cost}}{\partial b_u} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial b_u} + \frac{\lambda}{2} [2 * b_u] * \frac{\partial}{\partial b_u} \\ \frac{\partial \text{Cost}}{\partial b_m} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial b_m} + \frac{\lambda}{2} [2 * b_m] * \frac{\partial}{\partial b_m} \\ \frac{\partial \text{Cost}}{\partial \mu} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial \mu}\end{aligned}$$

## Task 2: Stochastic Gradient Descent

Now write down the update equations for each vector using stochastic gradient descent. Assume that you will use a constant step size ( $\alpha$ ):

$$\begin{aligned}\frac{\partial \text{Cost}}{\partial v_u} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial v_u} + \alpha * \frac{\lambda}{2} [\sum_u 2 * ||v_u||] * \frac{\partial}{\partial v_u} \\ \frac{\partial \text{Cost}}{\partial v_m} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial v_m} + \alpha * \frac{\lambda}{2} [\sum_m 2 * ||v_m||] * \frac{\partial}{\partial v_m} \\ \frac{\partial \text{Cost}}{\partial b_u} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial b_u} + \alpha * \frac{\lambda}{2} [2 * b_u] * \frac{\partial}{\partial b_u} \\ \frac{\partial \text{Cost}}{\partial b_m} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial b_m} + \alpha * \frac{\lambda}{2} [2 * b_m] * \frac{\partial}{\partial b_m} \\ \frac{\partial \text{Cost}}{\partial \mu} &= \sum_{(u,m) \in \text{Ratings}} 2 * \frac{1}{2} (\langle v_u, v_m \rangle + b_u + b_m + \mu - y_{u,m}) * \frac{\partial}{\partial \mu}\end{aligned}$$

Not sure if alpha term only corresponds to Penalty term