Internet Analytics (COM-308)

Homework Set 5

Exercise 1

The rating matrix of five users for four items with few missing entries is given by

$$R = \begin{bmatrix} 2 & 1 & 3 & - \\ 3 & 2 & 5 & 5 \\ 5 & - & 4 & 2 \\ 4 & 3 & - & 4 \\ - & 1 & 5 & 3 \end{bmatrix}.$$

(a) First we try a simple predictor that takes into account only user bias. Compute the optimal baseline predictor $\overline{r} + b_u$ (without regularization) and the RMSE for this predictor. Feel free to rely on a tool such as matlab, octave, or mathpy to solve the resulting system of equations.

(b) Now we refine this predictor to take into account both a user and an item bias, as seen in class. Compute the optimal baseline predictor $\overline{r} + b_u + b_i$ (without regularization) and the RMSE for this predictor.

(c) Predict the missing values in the rating matrix R for the predictors in parts (a) and (b).

Exercise 2

You are training a machine learning model on some training data, and then evaluate the error on some separate validation data that you kept aside. You notice that the validation error is considerably larger than the training error. Is this normal, or is there a problem? If there's a problem, how would you fix it by adjusting the regularizer weight λ? How will the two errors evolve if you change λ?

Exercise 3

(a) The most basic version of stochastic gradient descent (SGD) works as follows. There are n data points $(x_1, \ldots, x_n)$. We want to minimize a function $f(\theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(\theta; x_i)$ with respect to θ (which represents the model parameters to be optimized).

Instead of performing gradient descent using the full gradient $\nabla_\theta f$, we select, for each step in the iteration, a data index $I \sim \text{unif}(1, n)$ randomly, and we update the current estimate of θ with the gradient $\nabla_\theta f_I$.

What is the expected gradient $E[\nabla_\theta f_I]$?

(b) As we have seen in class, training the model on data is to solve an instance of the optimization problem above, with $x_i = (u, i, r_{ui})$, $\theta = (P, Q)$, and $f(\theta) = E(P, Q)$, i.e.,

$$(P^*, Q^*) = \arg\min_{P,Q} E(P, Q) = \arg\min_{P,Q} \sum_{(u,i)\in R} (r_{ui} - p_u^T q_i)^2 + \lambda(\|P\|^2 + \|Q\|^2).$$

Compute the full gradients $\nabla_P E(P, Q), \nabla_Q E(P, Q)$ used in gradient descent for this model.