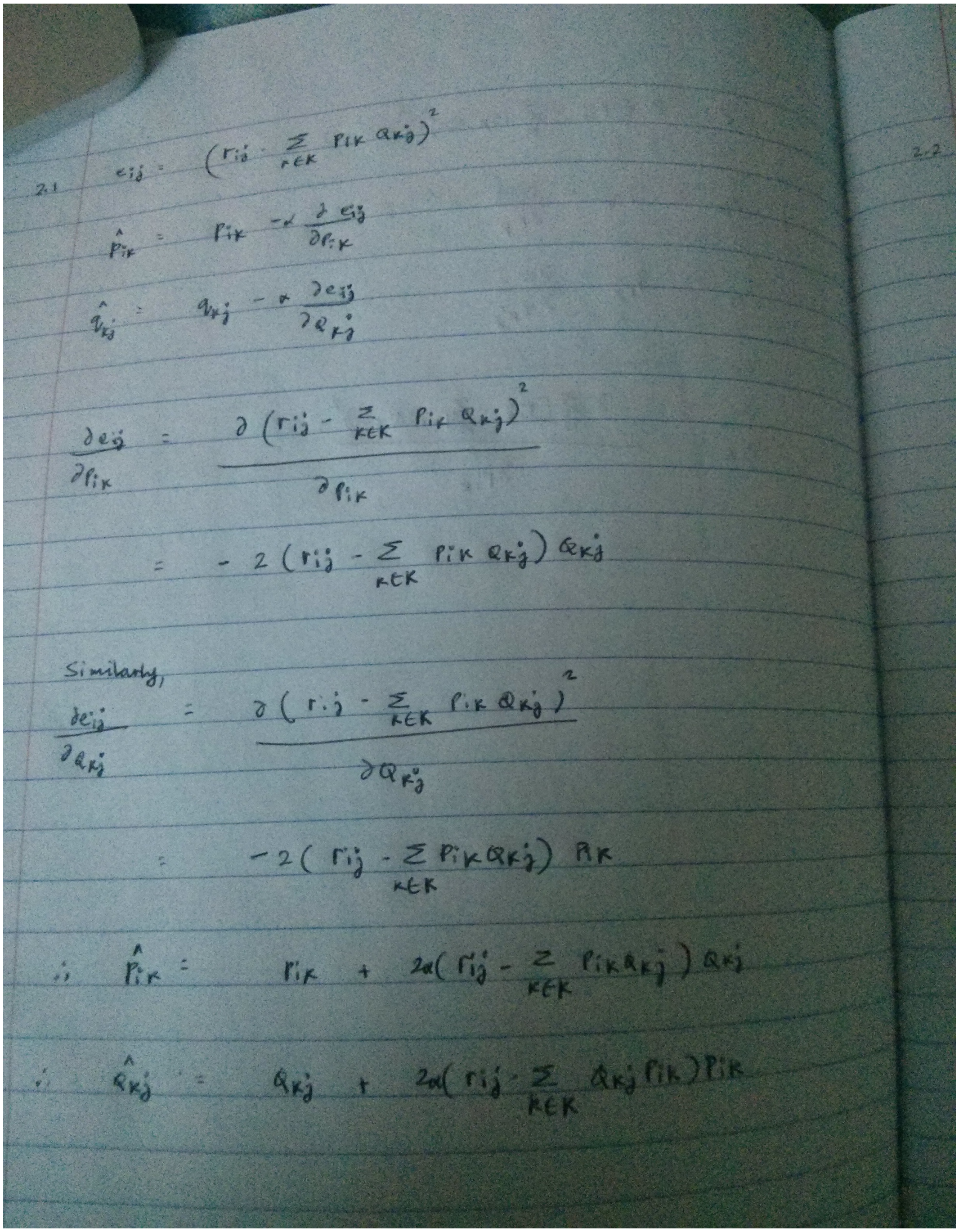
Web Search And Text Mining - A6
Anurag Kyal, 903040738

Derivation 2.1
Please note that $\alpha = \mu$ in my derivation


2.1 $$e_{ij} = \left(r_{ij} - \sum_{k \in K} p_{ik} q_{kj}\right)^2$$

$$\hat{p}_{ik} = p_{ik} - \alpha \frac{\partial e_{ij}}{\partial p_{ik}}$$

$$\hat{q}_{kj} = q_{kj} - \alpha \frac{\partial e_{ij}}{\partial q_{kj}}$$

$$\frac{\partial e_{ij}}{\partial p_{ik}} = \frac{\partial \left(r_{ij} - \sum_{k \in K} p_{ik} Q_{kj}\right)^2}{\partial p_{ik}}$$

$$= -2\left(r_{ij} - \sum_{k \in K} p_{ik} Q_{kj}\right) Q_{kj}$$

Similarly,

$$\frac{\partial e_{ij}}{\partial Q_{kj}} = \frac{\partial \left(r_{ij} - \sum_{k \in K} p_{ik} Q_{kj}\right)^2}{\partial Q_{kj}}$$

$$= -2\left(r_{ij} - \sum_{k \in K} P_{ik} Q_{kj}\right) P_{ik}$$

$$\therefore \hat{P}_{ik} = P_{ik} + 2\alpha\left(r_{ij} - \sum_{k \in K} P_{ik} Q_{kj}\right) Q_{kj}$$

$$\therefore \hat{Q}_{kj} = Q_{kj} + 2\alpha\left(r_{ij} - \sum_{k \in K} Q_{kj} P_{ik}\right) P_{ik}$$

Derivation 2.2

Please note that $\alpha == \mu$ and $\beta = \lambda$ in my derivation

2.2

$$e_{ij} = \Big(r_{ij} - \sum_{k \in K} p_{ik} q_{kj}\Big)^2 + \lambda\Big(\sum_{k \in K} p_{ik}^2 + \sum_{k \in K} q_{kj}^2\Big)$$

$$\frac{\partial e}{\partial p_{ik}} = -2\Big(r_{ij} - \sum_{k \in K} p_{ik} q_{kj}\Big) q_{kj} + 2\lambda p_{ik}$$

$$\frac{\partial e}{\partial q_{kj}} = -2\Big(r_{ij} - \sum_{k \in K} p_{ik} q_{kj}\Big) p_{ik} + 2\lambda q_{kj}$$

$$\hat{r}_{ik} = p_{ik} + 2\alpha\Big[\Big(r_{ij} - \sum_{k \in K} p_{ik} q_{kj}\Big) q_{kj} - \lambda p_{ik}\Big]$$

$$\hat{q}_{kj} = q_{kj} + 2\alpha\Big[\Big(r_{ij} - \sum_{k \in K} p_{ik} q_{kj}\Big) p_{ik} - \lambda q_{kj}\Big]$$

Cross Validation Tables

RMSE	$\lambda = 0.005$	$\lambda = 0.1$	$\lambda = 0.5$
$\mu = .0001$	0.9130	0.9139	0.9459
$\mu = .0005$	0.9134	0.9142	0.9461
$\mu = .001$	0.9136	0.9147	0.9467

Table 1: r = 1

RMSE	$\lambda = 0.005$	$\lambda = 0.1$	$\lambda = 0.5$
$\mu = .0001$	0.8958	0.8930	0.9354
$\mu = .0005$	0.9024	0.8949	0.9311
$\mu = .001$	0.9057	0.8969	0.9311

Table 2: r = 3

RMSE	$\lambda = 0.005$	$\lambda = 0.1$	$\lambda = 0.5$
$\mu = .0001$	0.9102	0.8960	0.9329
$\mu = .0005$	0.9196	0.9013	0.9301
$\mu = .001$	0.9220	0.9003	0.9305

Table 3: r = 5

μ	0.0001
λ	0.1
r	3
RMSE	0.8930

Table 4: r = 5

Analysis

1. What do you observe when you vary r? Why?

There is a variation in RMSE. For most combinations of λ and μ, RMSE decreases from r=3 and then increases for r=5 but is still smaller than for r=1.

Though intuitively it might seem that increasing the number of latent features(r), the results should improve, it is quite contrary. Gradual increase with r is justified but after a point, with further increase in r, the learner starts to overfit and thus performs poorly on the test data. Though experiment wasn't carried out to check the training error, it is expected to decrease with r.

2. Which model is the best? Please describe the best model in Table 4 and explain your choice.

The best model is which results in the lowest cross validation Root Mean Squared Error. This model in our combinations of parameters is with r=3, $\lambda = 0.1$ and $\mu = 0.0001$. The RMSE is 0.8930.

3. Suppose you are using regularized MF in real systems, how will you choose parameters? Why?

In real system with real data, I would carry out a similar experiment to calculate cross validation RMSE for different sets of parameters. I would chose a much larger suite of parameters and the final model will be the one with the lowest RMSE. Also, due to a larger set of data and large number of experiments, methods to parallelize the entire process will be useful.