- Foreword: I love this place indeed and therefore I shall work as hard as possible (like every hardworking Chinese) to stay, and make contributions.
- Collaborative filtering: Generally a form of supervised learning but with very special conditions.
- Classes of CF algorithms:
  - Memory based: Final prediction of rating typically obtained via a weighted sum of neighbors.
  - Model based: Based on constructing a model which describes important properties of the data.
- Association rule mining: Unable to generate recommendations for first time users, therefore it's user-based.
- Matrix factorization for CF: $R = P * Q$. Steps are as follows:
  1. Take, as input, matrix R, with elements r_{ij}.
  2. Create component matrices P and Q, by initializing randomly.
  3. Loop over all element of R which has been rated.
  4. Iterate until convergence.
- Regularized Matrix Factorization: By adjusting the prior distribution for the parameters, optimize the posterior distribution instead. (Reflected in years' exams. Try them tonight.)
- EM Algorithm:
  - Define full/complete data as $D = x, z$. x is the observed data, z is the hidden or missing data.
  - θ, the parameters of the model, also referred to as the complete data likelihood.

  In practice, we would like to optimize the log likelihood of the parameters based on the observed data. We can optimize this directly but in general it is very difficult to do so.

$$\theta_{ML} = argmax_\theta \log p(x;\theta)$$
$$\theta_{ML} = argmax_\theta \log \sum_z p(x,z;\theta)$$

- $$L(\theta) = \log \sum_z p(x,z;\theta)$$
$$L(\theta) = \log \sum_z Q(z) \frac{p(x,z;\theta)}{Q(z)}$$
$$L(\theta) \geq \sum_z Q(z) \log \frac{p(x,z;\theta)}{Q(z)}$$
- How do we make the bound tight? **we can make this tight if the expression in the log is a constant value**.
- $$\frac{p(x,z;\hat{\theta})}{Q(z)} = \frac{p(z|x;\hat{\theta})p(x;\hat{\theta})}{Q(z)}$$

  Therefore we set $Q(z) = p(z|x;\hat{\theta})$ where $\hat{\theta}$ is the current best guess of θ.
- – **Expectation:** Finding the distribution of z, given the data, and the current best guess of θ.
  – Calculating the expectation over the incomplete data likelihood over the current best guess of θ.

$$E_{\hat{\theta}}[p(x|\theta)] = \sum_z p(z|x;\hat{\theta}) \log \frac{p(x,z;\theta)}{p(z|x;\hat{\theta})} \quad (1)$$

  – **Maximization:** Optimize the equation with respect to θ.
- Some disadvantages for EM algorithm: it is still a maximum-likelihood approach, not Bayesian.
- In EM algorithm, what we are trying to maximize is the complete data log likelihood.
- Perceptron algorithm: it considers each training point in turn, adjusting the parameters to correct any mistakes. $\theta_{n+1} = \theta_n + y_i x_i$