Question 1

For display ads, we can use HTML text from the web pages which display those ads. Therefore, on the left hand side, we have HTML text as queries, and on the right hand side, we have display ads.

Question 2

The purpose of λ_2 and λ_3 is to prevent dividing by 0 when $R(i)$ or $R(u)$ is empty set. Another purpose is to diminish the impact to the biases when having small data sets.

Question 3

We can use k-fold cross-validation to estimate the performance of the model. Let's say we have 10 different latent factors, we train 10 individual models based on these 10 different latent factors and use k-fold cross-validation to select the best model.

Question 4

Rating matrix is a sparse matrix and does not fully use the information from those empty spaces. When integration the binary view of the rating matrix, we have a strong assumption that everybody watches every movies. A zero value in dense binary representation means that user does not care about that movie so that he or she leaves no rating. Therefore, dense binary representation gives us more information from the rating matrix and that produce the improvement.

Question 5

```
user: 15, movie: 7
rating with pearson correlation: 3.3
rating with content similarity: 2.3
user: 15, movie: 177
rating with pearson correlation: 2.6
rating with content similarity: 2.2
user: 19, movie: 921
rating with pearson correlation: 2.5
rating with content similarity: 3.8
user: 20, movie: 1089
rating with pearson correlation: 4.6
rating with content similarity: 3.7
user: 21, movie: 1393
rating with pearson correlation: 2.5
rating with content similarity: 3.3
```