

*질문1) Netflix Problem에서 사용했던 evaluation metric인 RMSE에 대해서 설명해주세요. 사용자 i 가 영화 j 에 대해 준 평점은 r_{ij} 라고 기입해주시고, 모델로 예측한 평점은 p_{ij} 일때, RMSE는 어떻게 계산할 수 있을까요? 전체 사용자 수는 N 명, 영화는 총 M 개 있다고 가정합니다.

$$\sqrt{\frac{1}{N} * \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^M (r_{ij} - p_{ij})^2}$$

- RMSE 수치와 실제 추천의 퀄리티와는 전혀 관련이 없다. RMSE를 아무리 낮게 맞추더라도 실제 Practical하게 사용하는 목적과 맞지 않을 수 있다. **추천을 어떻게 평가할 것이냐?** 라는 질문에 아직도 정해진 답은 없다. RMSE는 여러 평가 방법들 중 가장 간단한 평가 방법
- 실전에서는 특정 추천 로직을 배포한 다음 비교로직과의 지표(클릭, 구매율등 산업에서 중요하게 생각하는 것)를 기준으로 비교 평가를 한다.

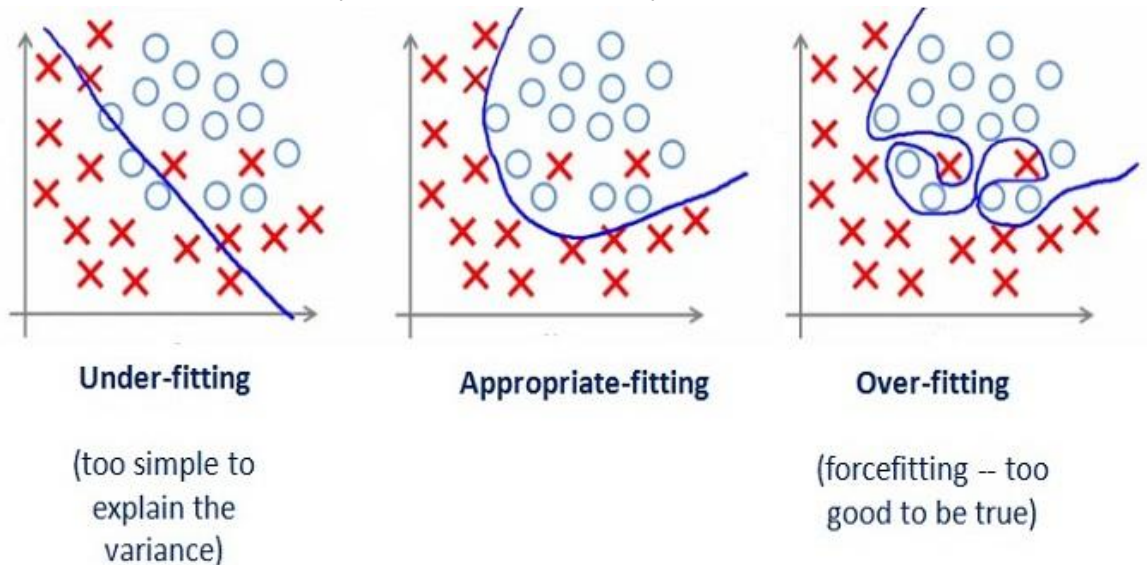
질문2) 문제 1에서 모델로 예측한 평점 p_{ij} 를 만드는 모델에는 여러 종류가 있습니다. 이 중에서 collaborative filtering 방법론 중 하나인 low rank matrix factorization은 p_{ij} 를 $p_{ij} = x_i^T y_j$ 로 만드는데요 (강의에서는 각각 x_i 와 θ_i 로 표현하며 x 와 y 는 모두 vector입니다) 이 방법이 왜 matrix factorization이라고 불리우는지 설명하고 x_i 와 y_j 의 의미에 대해서도 설명해주세요. (Hint: x_i 와 y_j 는 작은 차원의 vector이고 x_i 를 모아서 만든 행렬을 X , y_j 를 모아서 만든 행렬을 Y 라고 쓰면 전체 rating r_{ij} 를 Matrix로 나타낸 R 은 $R = X^T Y$ 로 표현이 가능합니다)

- 원본 rating 데이터를 Matrix로 표현한 후, 이 하나의 큰 matrix가 matrix 두 개의 곱으로 표현된다고 가정한다. 즉, $R = X^T Y$ 로 가정한다. 이때 R 은 N by M Matrix이고 X 는 N by d Y 는 M by d 로 가정한다. 만약 d 가 N 이나 M 에 비해 충분히 작다면 R 은 굉장히 적은 수의 차원으로 표현 가능한 행렬이 되는데 이를 개념적으로 "low rank" matrix라고 부른다. Low rank matrix factorization은 사용자의 평가 데이터를 Low rank matrix R 이라고 가정한 후, X 와 Y 라는 두개의 행렬로 분해 (factorization) 하기 때문에 low rank matrix factorization이라고 부른다.
- x_i 와 y_j 는 각각 사용자 i 와 영화 j 를 나타내는 d 차원 벡터가 된다. 사용자 i 가 영화 j 에 대해 가질 것으로 예측되는 선호도 p_{ij} 는 이 두 벡터의 내적 꼴로 표현할 수 있다.

질문3) Matrix Factorization 문제의 objective function은 $\sum_{ij} (r_{ij} - x_i^T y_j)^2 + \lambda (\sum_i \|x_i\|^2 + \sum_j \|y_j\|^2)$ 로 주어집니다 (λ 는 주어진 적당한 scalar 값, 예: 0.5 이고, $\|x\|$ 는 x 의 norm을 의미합니다. 단순히 생각하면 vector x 의 모든 element의 제곱합이라고 생각하시면 됩니다). 이 수식에서 앞 부분은 $p_{ij} = x_i^T y_j$ 이라는 hypothesis를 만족하기 위해 필요한 부분입니다. 그렇다면 뒷 부분은 어떤 이유로 필요할까요? 설명해주세요 (Hint: 다음 키워드와 연관되어있습니다. overfitting, regularization)

- λ 뒷부분은 Regularization 역할을 한다. 만약 이 텀이 없다면 모델이 자칫 overfitting 문제에 빠질 수 있기 때문이다. x 나 y 둘 중 하나를 고정하고 문제를 보면 입문 반에서 배운 Linear Regression과 정확히 같은 모양을 하고 있는 것을 알 수 있는데, Linear Regression에서 weight에 regularization 텀을 붙인 것과 같은 방식으로 이해할 수 있다.
- Objective function의 주 목적은 식을 최소화하는 것, 따라서 정규화 식의 x 와 y 도 작아야 한다. 그렇지만 마냥 작기만 한 식이 아니라 $R_{ij} - x_i^T y_j$ 도 최소화하는 식을 찾아야 한다. 비교적 튀지 않고 안정적인 값들이 나올 것이다. λ 를 키우면 키울수록 x 와 y 의 크기를 작게 만들게 강제효과가 있으니 안정적일 것이다.

λ 는 울퉁불퉁한 vector space를 부드러운 vector space로 조정한다.



- MF의 목적은 *평점을 예측* 하는거죠. 그런데 여기에서 overfitting이 발생하면 training 때 사용하지 않은 유저 - 아이템 페어에 대한 평점을 예측하면 원래 값에서 잘 예측이 안될 수 있습니다