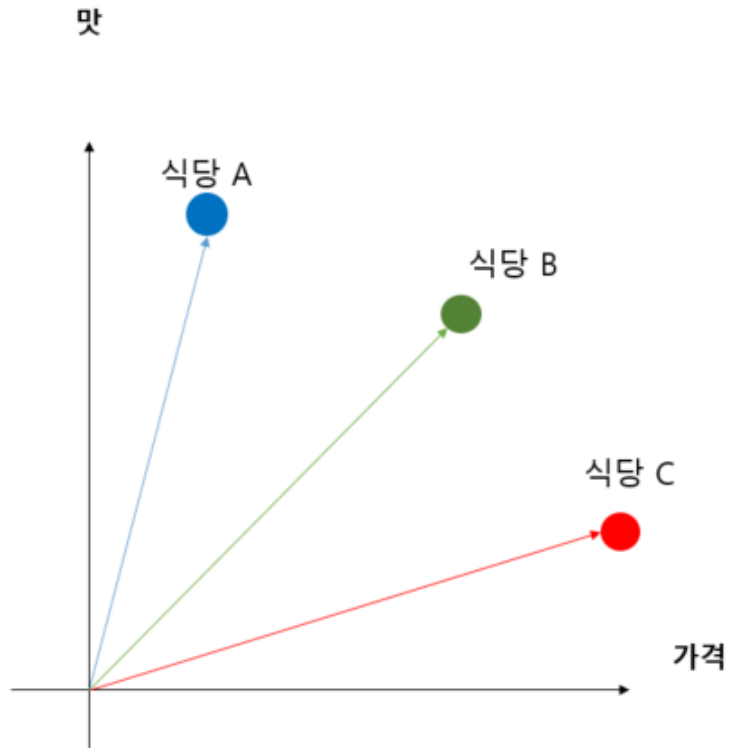


Recommend system

vector

- Vector: 크기 외에 방향이라는 성분을 가지고 있는 값
- 현실 세계의 많은 것들을 vector화 해서 vector space의 다양한 dimension으로 표현



- 식당 A는 가격은 싸고, 맛이 없는 곳.
- 식당 B는 가격은 비싸지만, 맛이 있는 곳.
- 식당 C는 가격이 비싸고, 맛이 없는 곳.

Vector similarity

- Cosine Similarity : 두 벡터 간의 코사인 각도를 이용하여 구할 수 있는 두 벡터의 유사도
- Vector Dot product : 두 벡터가 가까운 방향을 가지면 내적 값이 커지고 멀어질수록 작아진다.

$$similarity = \cos(\Theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

벡터의 내적 :

$$A \cdot B = a_1 * b_1 + a_2 * b_2 + \dots + a_n * b_n$$

```
doc1 = np.array([0,1,1,1])
doc2 = np.array([1,0,1,1])
doc3 = np.array([2,0,2,2])
```

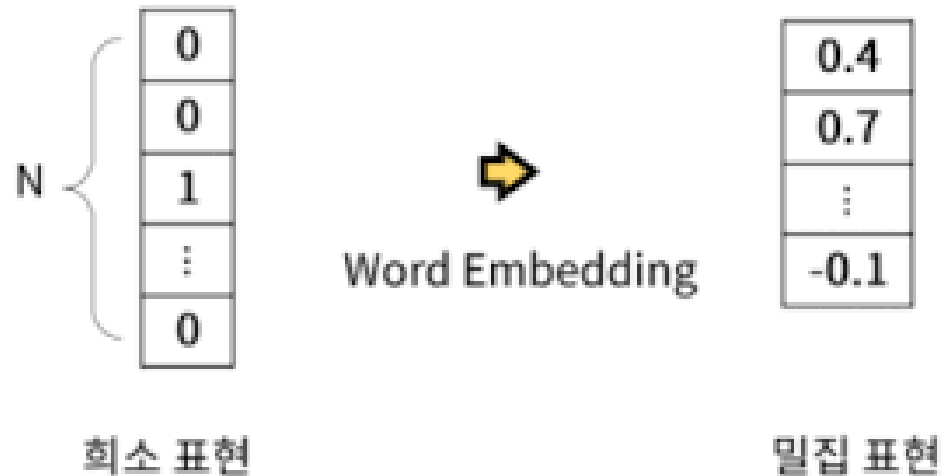
문서 1과 문서2의 유사도 : 0.67
문서 1과 문서3의 유사도 : 0.67
문서 2과 문서3의 유사도 : 1.00

코사인 거리 : 벡터의 크기에 영향을 받기 때문에, 벡터의 크기가 크게 다르면 유사도 측정이 왜곡될 수 있음

벡터의 내적: 벡터간의 내적과 벡터의 크기를 계산해야 함
내적보다 계산 cost가 높음

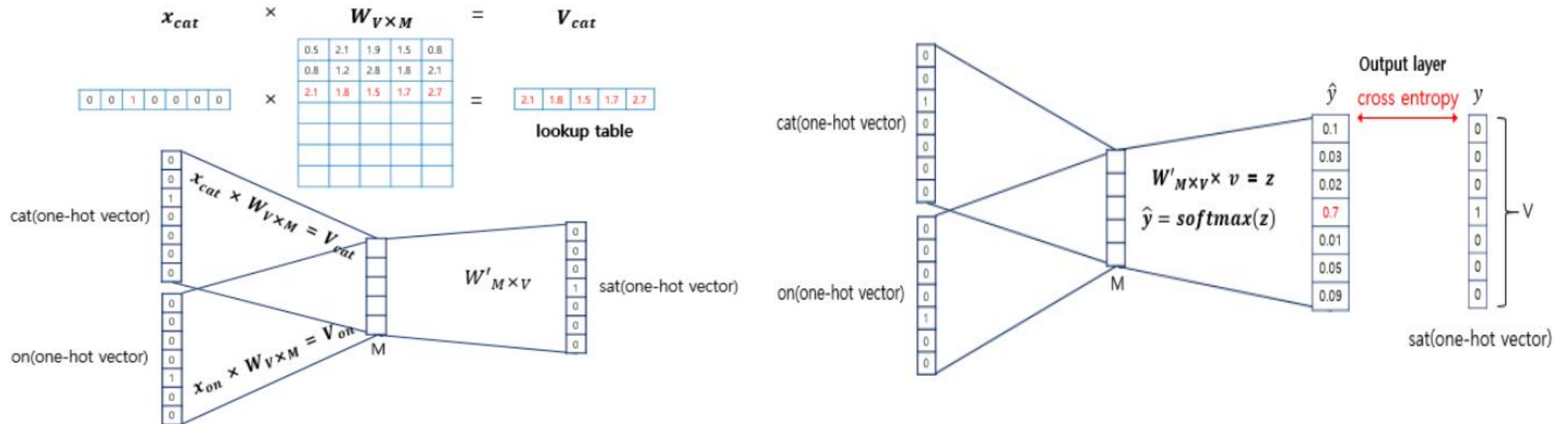
Embedding

- Embedding은 고차원에서 저차원의 벡터 공간으로 표현하는 기법
- Sparse(희소표현) ex) one hot encoding
- Dense(밀집표현) ex) word to vector



word2vec

- CBOW(Continuous Bag of Words) : 주변단어들의 입력으로 중간에 있는 단어 예측
- Skip-Gram : 중간에 있는 단어들의 입력으로 주변단어들을 예측



추천시스템

- 콘텐츠 기반 필터링 모델(Content-based Filtering)
- 협업 필터링 모델(Collaborative Filtering, CF)
 - - 메모리기반
 - - 모델 기반
 - - 융합모델

메모리 기반 접근 방식

- - 사용자 기반 추천 (User-based Recommendation)
 - - 아이템 기반 추천 (Item-based Recommendation)
-
- 유사도를 기반으로 동작
 - 코사인 유사도 : 두 벡터가 가리키는 방향이 얼마나 유사한가를 의미
 - 피어슨 유사도 : 피어슨 유사도가 1이면 양의 상관관계(비교하는 데이터 중 하나가 증가하면 다른 하나도 증가함), -1이면 음의 상관관계(하나가 증가하면 다른 하나는 감소함), 0이면 상관관계가 없음(독립)을 의미

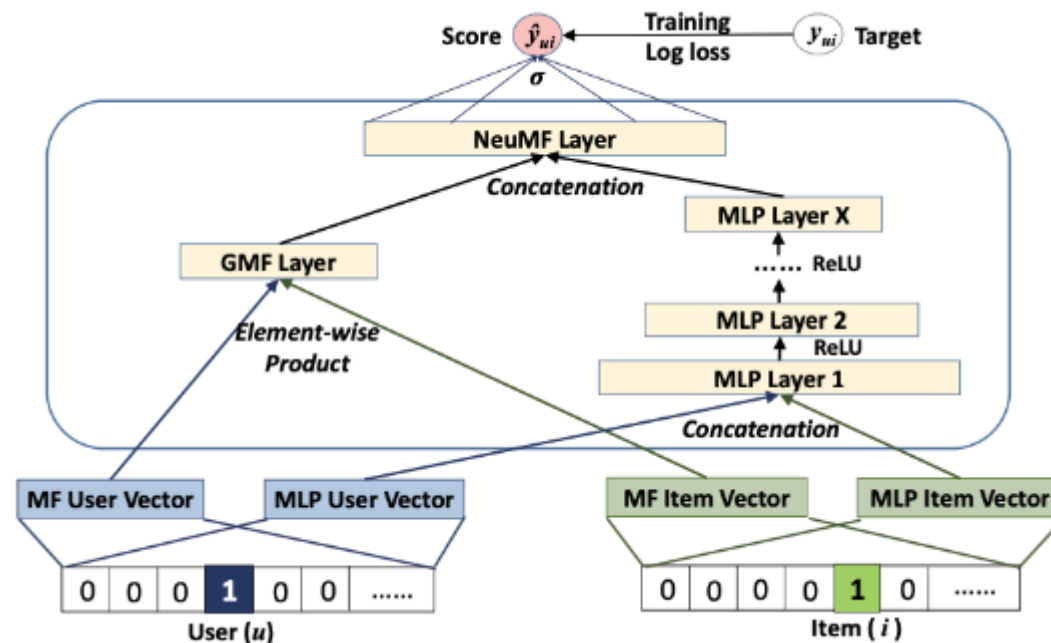
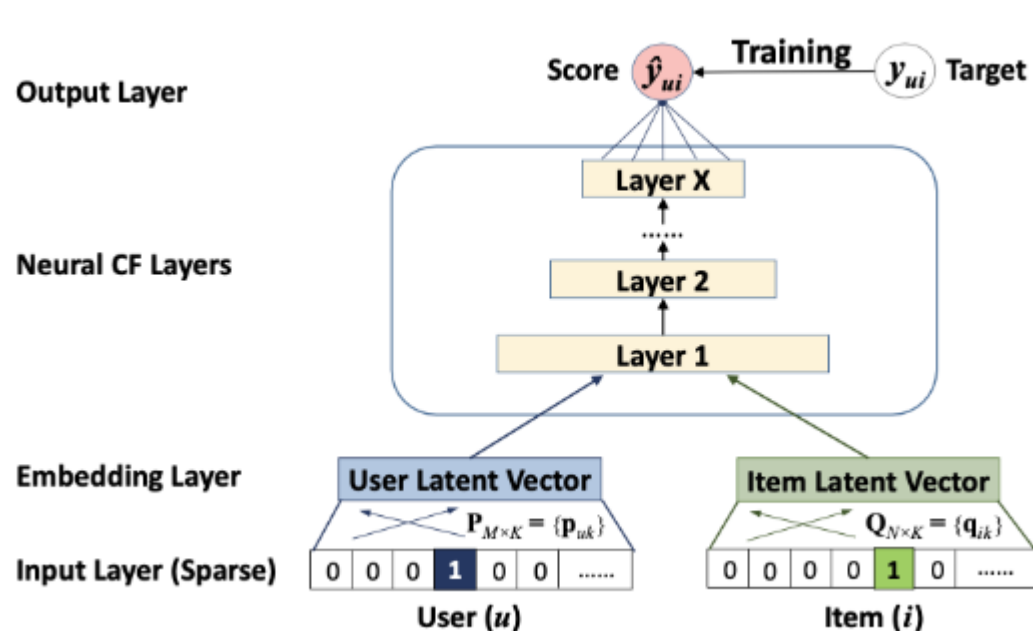
모델 기반 접근방식

- Latent Factor 모델
- Matrix factorization
- Matrix를 랜덤값으로 채움
- 사용자 벡터와 아이템 벡터를 내적하면 요소값(평점)이 나올것이다.“ 라고 가정한 뒤 $(0.3, 0.4, 0.4) * (0.2, 0.3, 0.7) = 4.6$
- 사용자 벡터와 아이템 벡터의 내적값(예상평점)이 실제평점값과 유사해지도록 두 개의 Matrix를 학습 시키면 된다

Item Matrix P			User Matrix Q ^T											
0.3	0.4	0.4		4		5			5	4.2		3		1
0.4	0.4	0.1	3	1	2			4			4	5		
0.4	0.2	0.8		5	3	4		3		2	1		4	2
0.7	0.5	0.9		2			4			5		4	2	
0.1	0.1	0.6	5	2					2	4	3	4		
0.1	0.3	0.1		4			2			3		3		1

Neural Collaborative Filtering

■ Neural Collaborative Filtering



- user-item간의 상호 관계를 표현하기 위해 MF의 linearity 와 MLP의 non-linearity를 결합한 것이 특징
- 이 모델을 neural matrix factorization 함