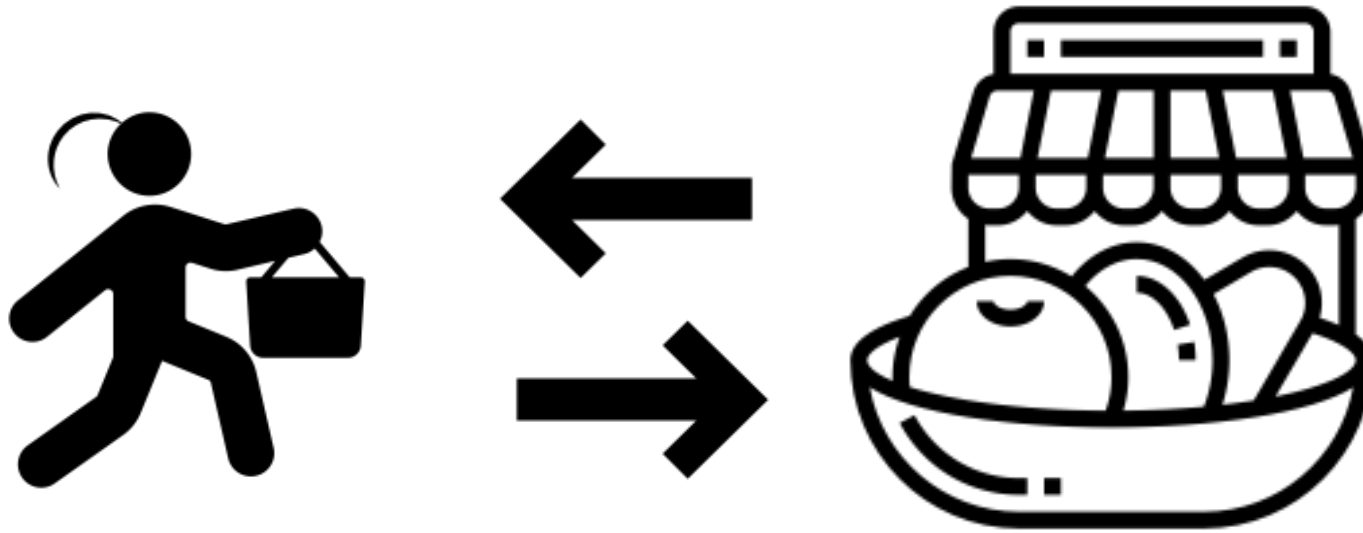


추천 시스템 모형 평가

추천 시스템의 목적

핵심 이슈 : 고객에게 어떠한 제품을 권할 것인가?

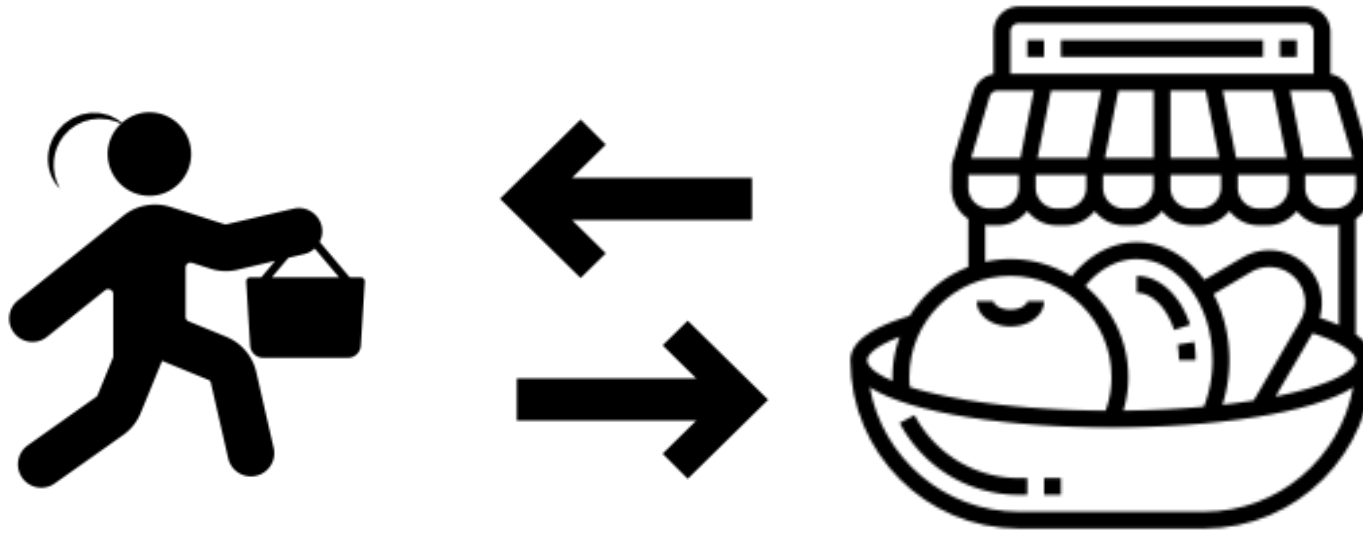


고객

상점

추천 시스템의 목적

핵심 이슈 : 고객에게 **어떠한** 제품을 권할 것인가?



고객

상점

기업에서 추천 시스템을 만드는 목적

기업의 **매출**, **이윤** 증대

좋은 추천 시스템의 기준

주요 정량 평가 기준들

CTR(Click Through Rate)

$$CTR = \frac{\textit{click}}{\textit{impressions}} \quad : \text{노출 대비 특정 링크를 클릭한 유저의 비율}$$

VTR(View Through Rate)

$$VTR = \frac{\textit{View Through}}{\textit{impressions}} \quad : \text{노출 대비 조회한 유저의 비율}$$

CVR(Conversion Rate)

$$CVR = \frac{\textit{Conversion}}{\textit{impressions}} \quad : \text{노출 대비 회원가입, 구매한 유저의 비율}$$

좋은 추천 시스템의 기준

주요 정량 평가 기준들

CTR(Click Through Rate)

$$CTR = \frac{click}{impressions} \quad : \text{노출 대비 특정 링크를 클릭한 유저의 비율}$$

VTR(View Through Rate)

$$VTR = \frac{View Through}{impressions} \quad : \text{노출 대비 조회한 유저의 비율}$$

CVR(Conversion Rate)

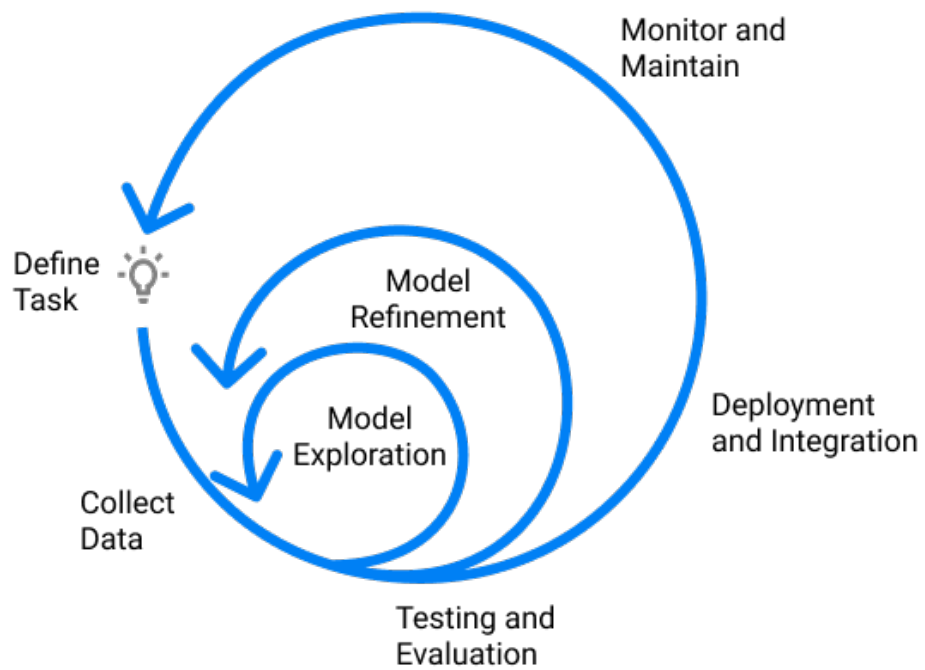
$$CVR = \frac{Conversion}{impressions} \quad : \text{노출 대비 회원가입, 구매한 유저의 비율}$$

CTR / VTR / CVR은 추천 시스템의 성능을 평가하는 핵심 기준

머신러닝 모델의 개발 과정

Trial & Error 방식으로, 많은 실험을 반복해가며 모델을 개발해야 함

Machine Learning Development Lifecycle



추천 시스템 평가 방식

1. Offline Evaluation

과거 Interaction 기록을 바탕으로 모델을 평가



2. Online Evaluation

모델을 배포하여 고객들의 실제 피드백을 바탕으로 평가 (A/B Test)

A/B Test : 기존 모델(A)과 변형 모델(B)을 동시에 고객에게 노출하여 평가하는 종합 대조 실험



Offline 평가 지표

Precision@K

추천 아이템 K개 중 실제로 고객이 구매한 아이템의 수

MRR@K (Mean Reciprocal Rank)

<추천되어야 하는 상품 K개의 순위가 얼마나 밀렸는가?>

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank_i}$$

NDCG@K (Normalized Discounted Cumulative Gain)

이상적인 추천 결과로부터 구할 수 있는 최대 가능 스코어를 기준으로 정규화한, 관련 아이템 순위에 역로그 가중치를 반영하여 계산

$$nDCG(k, V_U) = \frac{1}{|V_U|} \sum_{u \in V_U} \frac{1}{Z_u} \sum_{i=1}^k \frac{2^{r^u} - 1}{\log(1 + i)}$$

검색시스템의 평가 기준과 유사

AP@K (Average Precision)

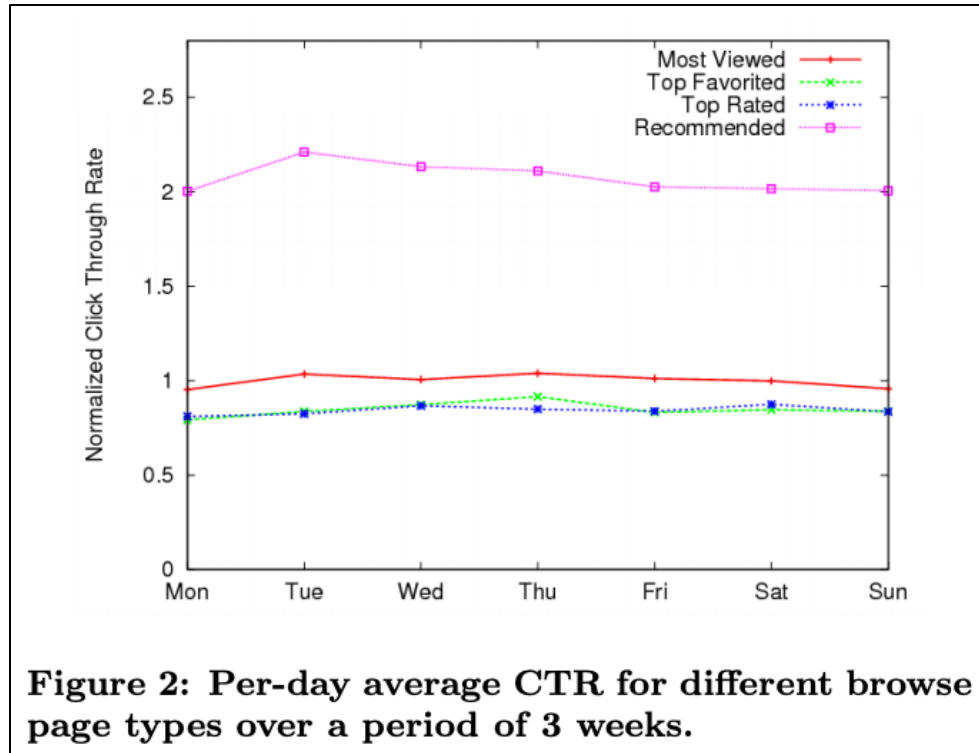
모든 관련 아이템의 순위 중 Top-K 아이템의 평균 정밀도

Offline 평가 지표

Coverage	<p>In the recommender system, the coverage of items refers to the proportion of items recommended to total items [26]. The coverage of items is defined as follows:</p> $coverage = \frac{\bigcup_{u \in U} I(u)}{I} \tag{13}$ <p>where $I(u)$ is the number of items recommended for a user, I represents the total number of items. Coverage is an important evaluation metric of recommender systems because it can describe the ability of mining the range of recommendation items. A good recommendation system requires not only high prediction accuracy but also high coverage.</p>
Diversity	<p>In Recommender systems, the definition of the diversity of items is just opposite to that of similarity. In some situations, it makes no sense to recommend similar items for the users in practice. For example, if the user has already bought a sport watch with GPS function, generally speaking, he will not be interested in other similar sport watches from other brands recommended by the recommender system. But if the recommender system can recommend him some other items, like the heart rate belt or sport music CD, it can make much better recommendation results. Hence, while designing a recommender system, not only the accuracy of prediction, but also the diversity of recommendation products should be concerned, so as to satisfy different requirements of users [30].</p>
Serendipity	<p>In a recommender system, the serendipity means the serendipitous levels that the system can bring about [31]. For example, if a user like the films stared by a certain film star, then he will feel like it or have some novelty when the system recommends him some early films stared by the film star, but he will not feel surprised or unexpected. A random recommendation will bring serendipity to the users, but if too many unrelated items are recommended, the trust on the recommender system will be reduced. Therefore, it is becoming more and more concerning to the researchers on how to balance prediction accuracy and serendipity. To evaluate serendipity levels, first the similarity between the recommended items and the previous preferred items of the user should be defined, and second, the satisfaction of the users on the recommended items should be statistically recorded. The first step can be calculated through offline analytics and the second step should be accomplished by users' survey.</p>

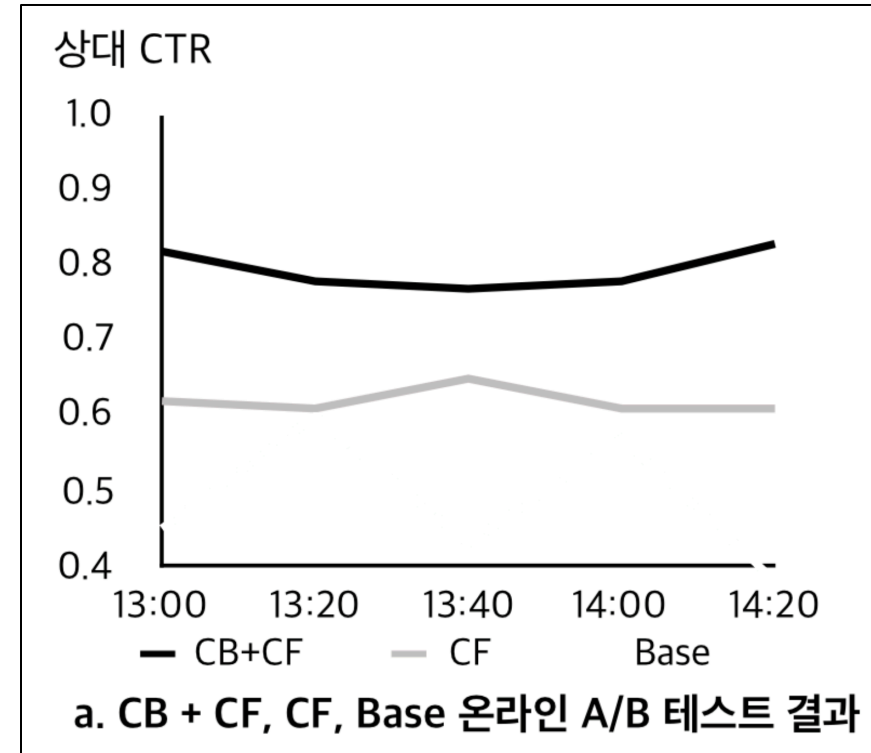
Online 평가 방식 예)

1. Youtube Recommendation System



Youtube Recommendation paper, 2010

2. Kakao Recommendation System



카카오 추천시스템 관련 글

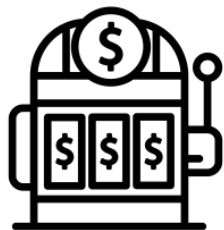
<https://brunch.co.kr/@kakao-it/72>

Online 평가의 최적화 문제, Multi Armed Bandit

문제 상황

아래와 같이 3가지 모형 TOP-K, Item-based CF, BPR 알고리즘을 바탕으로 구성된 추천 엔진이 있다고 하자. 매출을 최대화하기 위해서는 3가지 모형 중 CTR(Click Through Rate)가 가장 높은 모형을 찾아야 함
하지만 우리는 각 모형의 CTR을 모르고 있을 때, 우리는 어떻게 해야 할까?

Top-K



Item-based CF



BPR



CTR

?

?

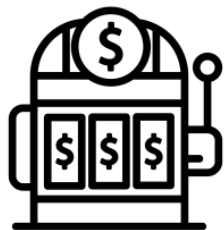
?

Online 평가의 최적화 문제, Multi Armed Bandit

문제 상황

아래와 같이 3가지 모형 TOP-K, Item-based CF, BPR 알고리즘을 바탕으로 구성된 추천 엔진이 있다고 하자. 매출을 최대화하기 위해서는 3가지 모형 중 CTR(Click Through Rate)가 가장 높은 모형을 찾아야 함
하지만 우리는 각 모형의 CTR을 모르고 있을 때, 우리는 어떻게 해야 할까?

Top-K



Item-based CF



BPR



우리의 목적

고객을 최대한 덜 놓치면서, 제일 좋은 모형을 찾는 것

CTR

?

?

?

Multi Armed Bandit의 알고리즘 (1) Epsilon Greedy

가장 간단한 형태의 알고리즘 탐색 방법

A simple bandit algorithm

Initialize, for $a = 1$ to k :

$Q(a) \leftarrow 0$

$N(a) \leftarrow 0$

Repeat forever:

$A \leftarrow \begin{cases} \arg \max_a Q(a) & \text{with probability } 1 - \varepsilon \quad (\text{breaking ties randomly}) \\ \text{a random action} & \text{with probability } \varepsilon \end{cases}$

$R \leftarrow \text{bandit}(A)$

$N(A) \leftarrow N(A) + 1$

$Q(A) \leftarrow Q(A) + \frac{1}{N(A)} [R - Q(A)]$

Top-K Item-based CFBPR



CTR

?

?

?

Epsilon Greedy

탐색(Exploration)과 활용(Exploit)을 나누어서 진행하는 방식. epsilon(ex : 10%)만큼은 탐색에 이용.

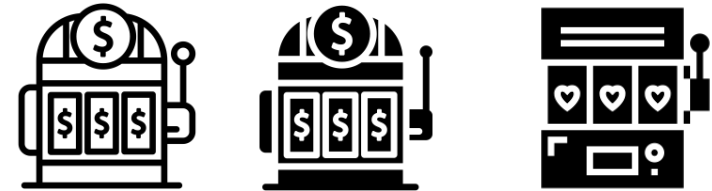
- 탐색(exploration) : 무작위로 모델을 이용해 추천해보며 정보가 부족한 모델의 CTR을 유추
- 활용(exploit) : 지금까지 관측된 슬롯머신 중 승률이 제일 좋은 것만 시도

Multi Armed Bandit의 알고리즘 (2) Thomson Sampling

현재 가장 일반적으로 쓰이는 방식

$$\begin{aligned}\hat{\theta}_k &\sim \text{Beta}(\alpha_k, \beta_k) \\ \theta &= \max(\hat{\theta}_k) \\ k &= \operatorname{argmax}(\hat{\theta}_k) \\ \text{reward}_t &\sim \text{Bernoulli}(p = \theta) \\ \alpha_k, \beta_k &= \alpha_k + \text{reward}, \beta_k + 1 - \text{reward}\end{aligned}$$

Top-K Item-based CFBPR



CTR

?

?

?

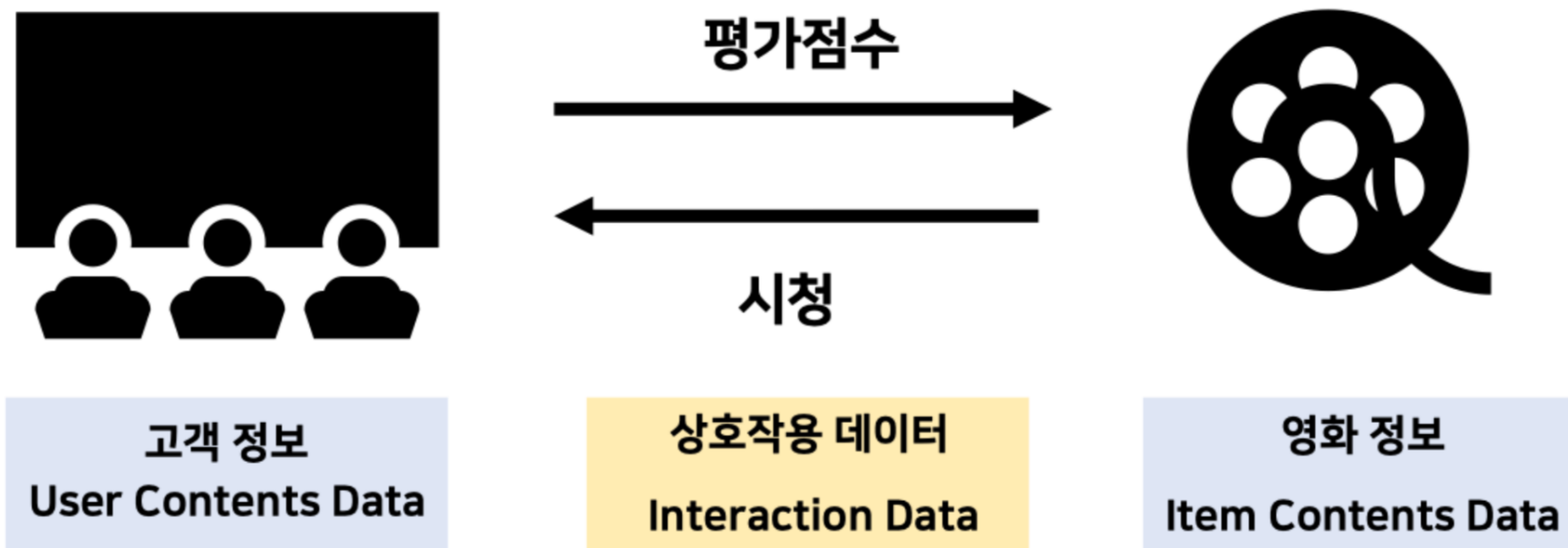
Thompson Sampling

탐색(Exploration)과 활용(Exploit)을 확률을 근간해 수행하는 방법

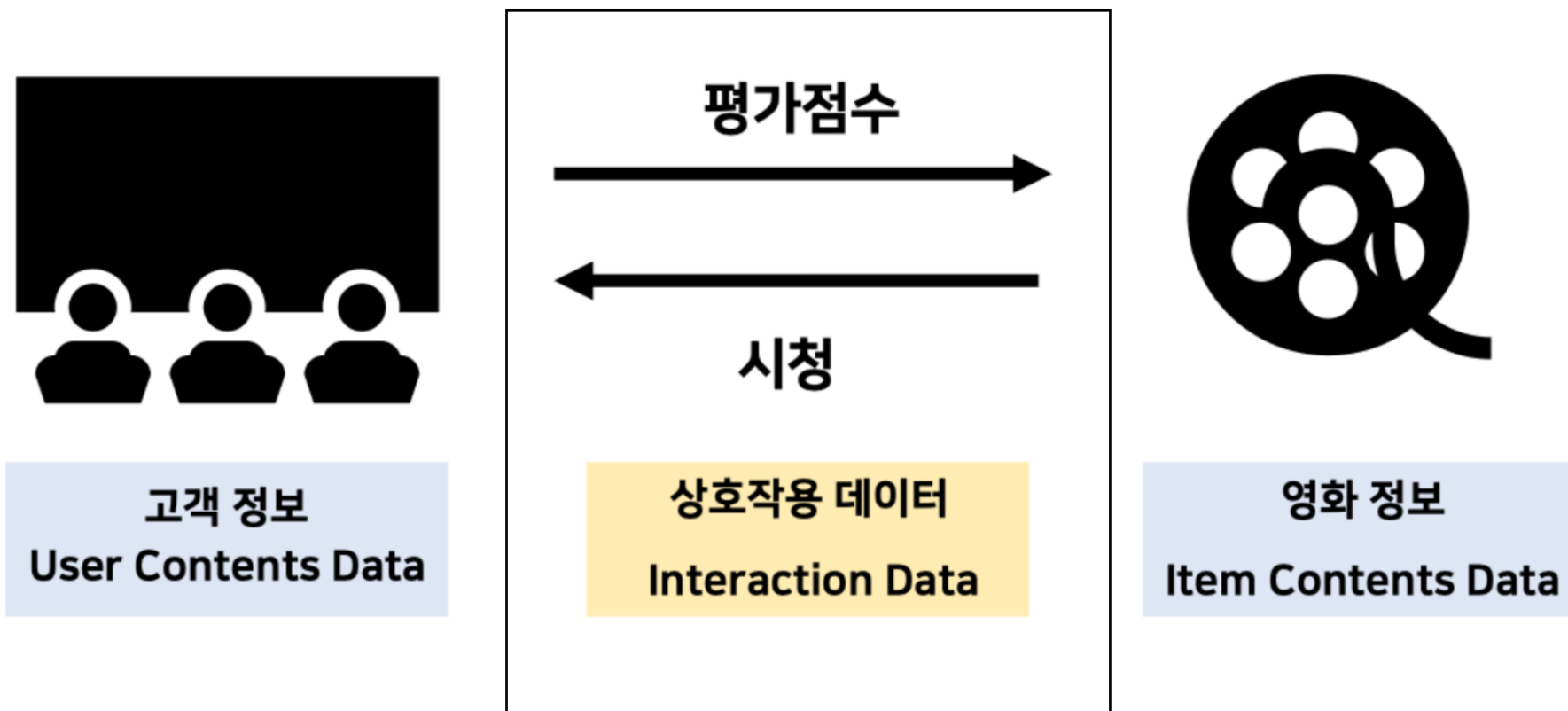
- 베타 분포(Beta Distribution)를 활용해 CTR의 분포를 추론
- 시도가 반복됨에 따라, CTR의 분포는 특정값에 수렴

Factorization Machine

이번 시간부터 다룰 데이터

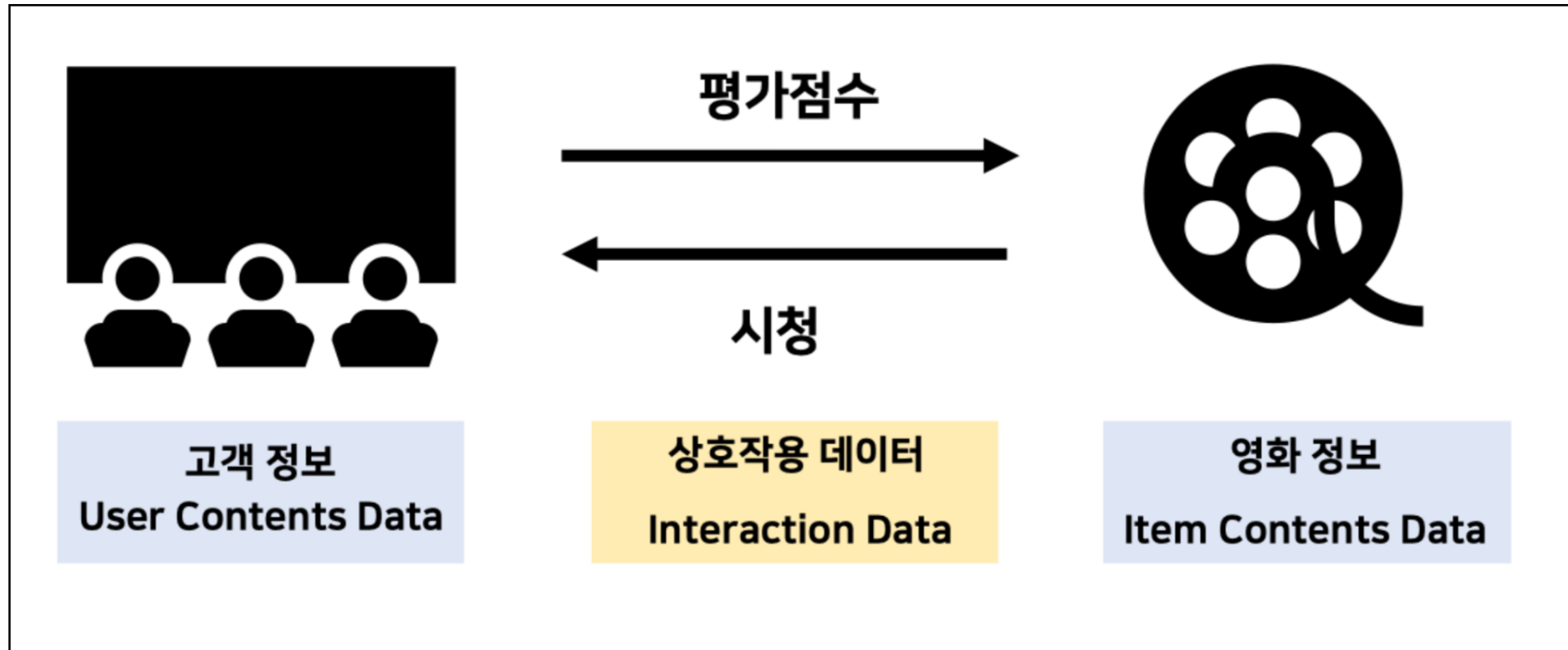


이번 시간부터 다룬 데이터



CF 알고리즘이 다룬 데이터 범위

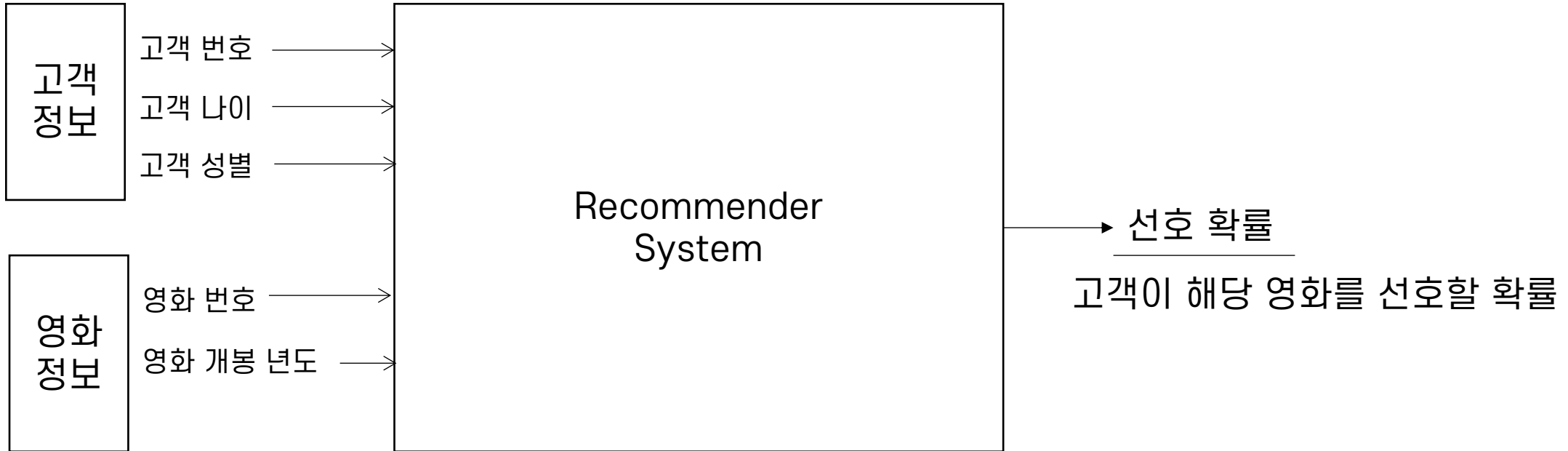
이번 시간부터 다룰 데이터



이번시간부터 다룰 Factorization Machine의 데이터 범위

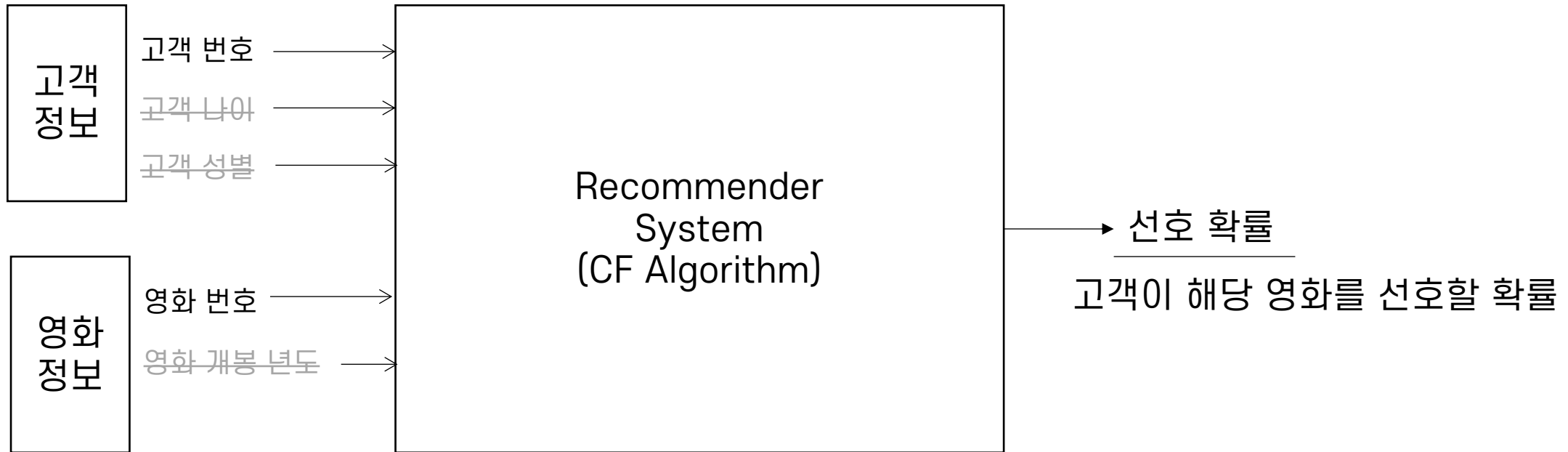
우리의 목표

고객 정보와 영화 정보를 활용하여 보다 정확한 추천을 할 수 있는 모델을 개발



우리의 목표

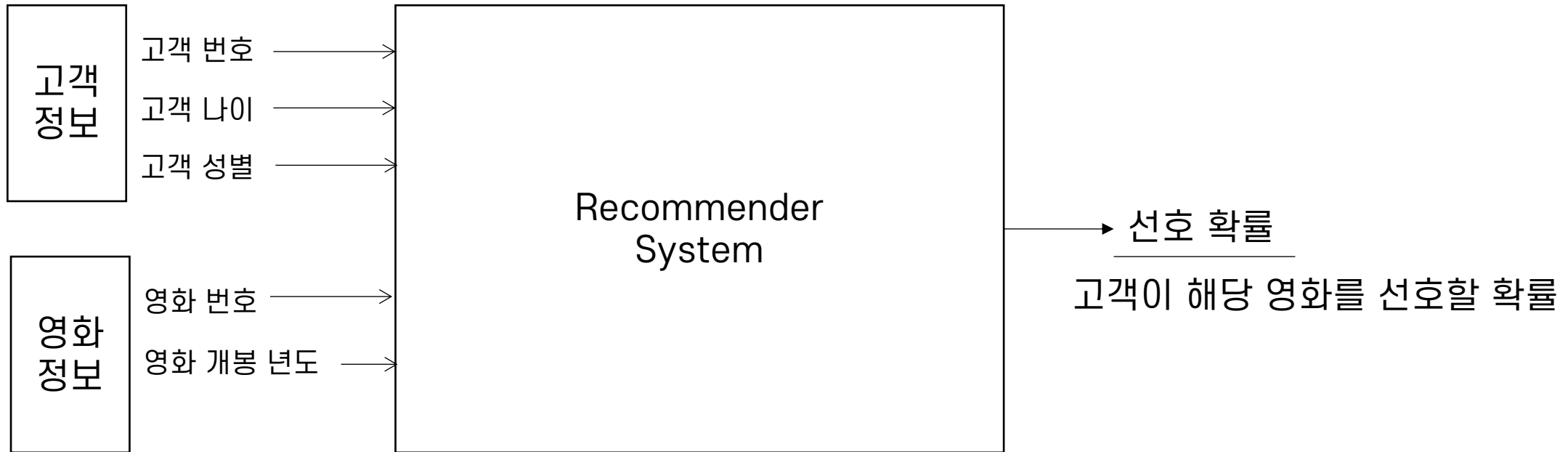
고객 정보와 영화 정보를 활용하여 보다 정확한 추천을 할 수 있는 모델을 개발



고객 번호와 영화 번호만으로 선호 확률을 예측한 것 = Collaborative Filtering

가장 간단한 형태의 추천시스템 모형 : 선형회귀

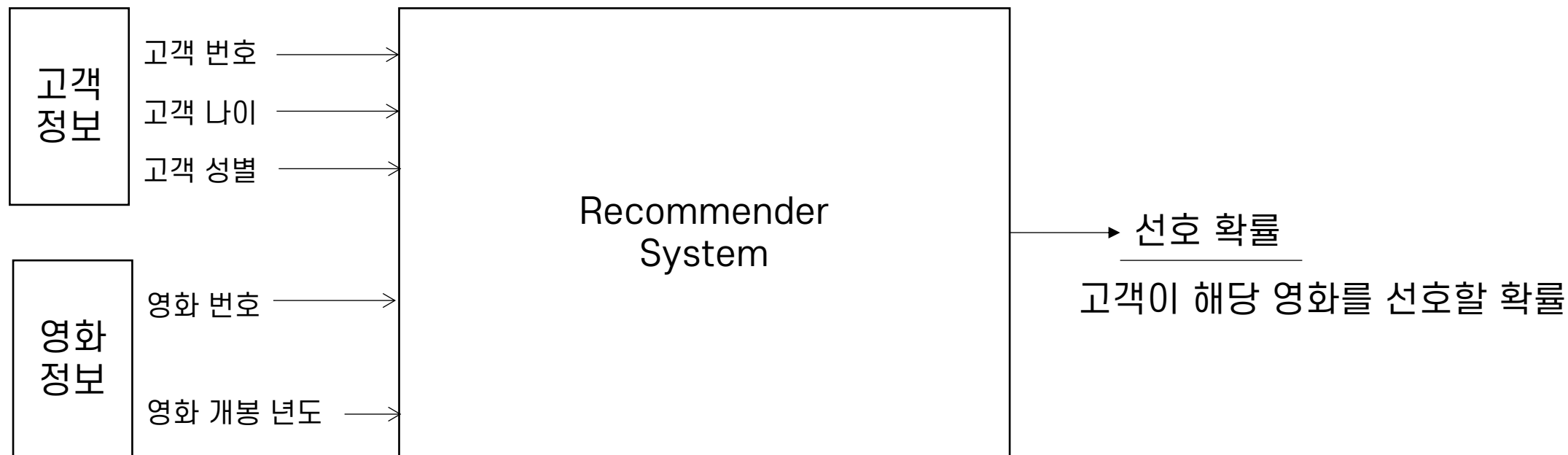
고객 정보와 영화 정보를 활용하여 보다 정확한 추천을 할 수 있는 모형을 개발



$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

가장 간단한 형태의 추천시스템 모형 : 선형회귀

고객 정보와 영화 정보를 활용하여 보다 정확한 추천을 할 수 있는 모형을 개발



$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

➤ 여기서 입력값들은 모두 범주형 데이터

가장 간단한 형태의 추천시스템 모형 : 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

가장 간단한 형태의 추천시스템 모형 : 선형회귀

$$y = \sigma(W_{\text{고객번호}} \underline{X_{\text{고객번호}}} + W_{\text{나이}} \underline{X_{\text{나이}}} + W_{\text{성별}} \underline{X_{\text{성별}}} + W_{\text{영화번호}} \underline{X_{\text{영화번호}}} + W_{\text{연도}} \underline{X_{\text{연도}}} + W_0)$$

입력 변수 : 범주형 데이터(Category)

가장 간단한 형태의 추천시스템 모형 : 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

각 범주 별 값들(클래스)

- 고객번호 : 1번, 2번, 3번, ..., 943번
- 나이 : (0~4세), (5~9세), (10~14세), (15~19세), (20~24세), ..., (65~69세)
- 성별 : 남 / 여
- 영화번호 : 1번, 2번, 3번, ..., 1682번
- 연도 : 1922년 ~ 1998년

가장 간단한 형태의 추천시스템 모형 : 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

각 범주 별 값들(클래스)

- 고객번호 : 1번, 2번, 3번, ..., 943번
 - 나이 : (0~4세), (5~9세), (10~14세), (15~19세), (20~24세), ..., (65~69세)
 - 성별 : 남 / 여
 - 영화번호 : 1번, 2번, 3번, ..., 1682번
 - 연도 : 1922년 ~ 1998년
-

이것을 어떻게 숫자로 표현할까?

가장 간단한 형태의 추천시스템 모형 : 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

각 카테고리 별 값들(클래스)

- 고객번호 : 1번, 2번, 3번, ..., 943번
 - 나이 : (0~4세), (5~9세), (10~14세), (15~19세), (20~24세), ..., (65~69세)
 - 성별 : 남 / 여
 - 영화번호 : 1번, 2번, 3번, ..., 1682번
 - 연도 : 1922년 ~ 1998년
-

이것을 어떻게 숫자로 표현할까?

-> one-hot encoding

one-hot Encoding이란?

명목형 변수를 숫자로 바꾸는 방법으로,
해당 클래스를 지칭하는 인덱스에만 1을 두고, 그 외에는 모두 0으로 두는 방식

```
sample_df = data_df.sample(5, random_state=0)  
sample_df.occupation
```

```
59041      other  
50627  educator  
82548      other  
65372  programmer  
15823      student  
Name: occupation, dtype: object
```

```
pd.get_dummies(sample_df.occupation)
```

	educator	other	programmer	student
59041	0	1	0	0
50627	1	0	0	0
82548	0	1	0	0
65372	0	0	1	0
15823	0	0	0	1

one-hot Encoding의 이점

예시) 직업이라는 데이터를 이용해 수식 전개할 경우

```
sample_df = data_df.sample(5, random_state=0)
sample_df.occupation
```

```
59041      other
50627  educator
82548      other
65372  programmer
15823    student
Name: occupation, dtype: object
```

```
pd.get_dummies(sample_df.occupation)
```

	educator	other	programmer	student
59041	0	1	0	0
50627	1	0	0	0
82548	0	1	0	0
65372	0	0	1	0
15823	0	0	0	1

$$y_0 = w_0 + w_{educator} * 0 + w_{other} * 1 + w_{programmer} * 0 + w_{student} * 0$$

$$y_1 = w_0 + w_{educator} * 1 + w_{other} * 0 + w_{programmer} * 0 + w_{student} * 0$$

$$y_2 = w_0 + w_{educator} * 0 + w_{other} * 1 + w_{programmer} * 0 + w_{student} * 0$$

$$y_3 = w_0 + w_{educator} * 0 + w_{other} * 0 + w_{programmer} * 1 + w_{student} * 0$$

$$y_4 = w_0 + w_{educator} * 0 + w_{other} * 0 + w_{programmer} * 0 + w_{student} * 1$$

명목형 변수를 위한 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

명목형 변수를 위한 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

↓ 원핫 인코딩으로 변형

$$y = \sigma(\\ w_{\text{고객번호}_1\text{번}} x_{\text{고객번호}_1\text{번}} + w_{\text{고객번호}_2\text{번}} x_{\text{고객번호}_2\text{번}} + \cdots + w_{\text{고객번호}_{973}\text{번}} x_{\text{고객번호}_{973}\text{번}} + \\ w_{0\sim4\text{세}} x_{0\sim4\text{세}} + w_{5\sim9\text{세}} x_{5\sim9\text{세}} + w_{10\sim14\text{세}} x_{10\sim14\text{세}} + \cdots + w_{65\sim69\text{세}} x_{65\sim69\text{세}} + \\ w_{\text{남}} x_{\text{남}} + w_{\text{여}} x_{\text{여}} + \\ w_{\text{영화번호}_1\text{번}} x_{\text{영화번호}_1\text{번}} + w_{\text{영화번호}_2\text{번}} x_{\text{영화번호}_2\text{번}} + \cdots + w_{\text{영화번호}_{1682}\text{번}} x_{\text{영화번호}_{1682}\text{번}} + \\ w_{1922\text{년}} x_{1922\text{년}} + w_{1923\text{년}} x_{1923\text{년}} + \cdots + w_{1998\text{년}} x_{1998\text{년}})$$

명목형 변수를 위한 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

↓ 원핫 인코딩으로 변형

$$y = \sigma(\\ w_{\text{고객번호}_1\text{번}} x_{\text{고객번호}_1\text{번}} + w_{\text{고객번호}_2\text{번}} x_{\text{고객번호}_2\text{번}} + \cdots + w_{\text{고객번호}_{973}\text{번}} x_{\text{고객번호}_{973}\text{번}} + \\ w_{0\sim4\text{세}} x_{0\sim4\text{세}} + w_{5\sim9\text{세}} x_{5\sim9\text{세}} + w_{10\sim14\text{세}} x_{10\sim14\text{세}} + \cdots + w_{65\sim69\text{세}} x_{65\sim69\text{세}} + \\ w_{\text{남}} x_{\text{남}} + w_{\text{여}} x_{\text{여}} + \\ w_{\text{영화번호}_1\text{번}} x_{\text{영화번호}_1\text{번}} + w_{\text{영화번호}_2\text{번}} x_{\text{영화번호}_2\text{번}} + \cdots + w_{\text{영화번호}_{1682}\text{번}} x_{\text{영화번호}_{1682}\text{번}} + \\ w_{1922\text{년}} x_{1922\text{년}} + w_{1923\text{년}} x_{1923\text{년}} + \cdots + w_{1998\text{년}} x_{1998\text{년}})$$

Q) 15세 남자인 고객번호 4번이 1923년에 개봉한 영화번호 2번을 선호할 확률은?

명목형 변수를 위한 선형회귀

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

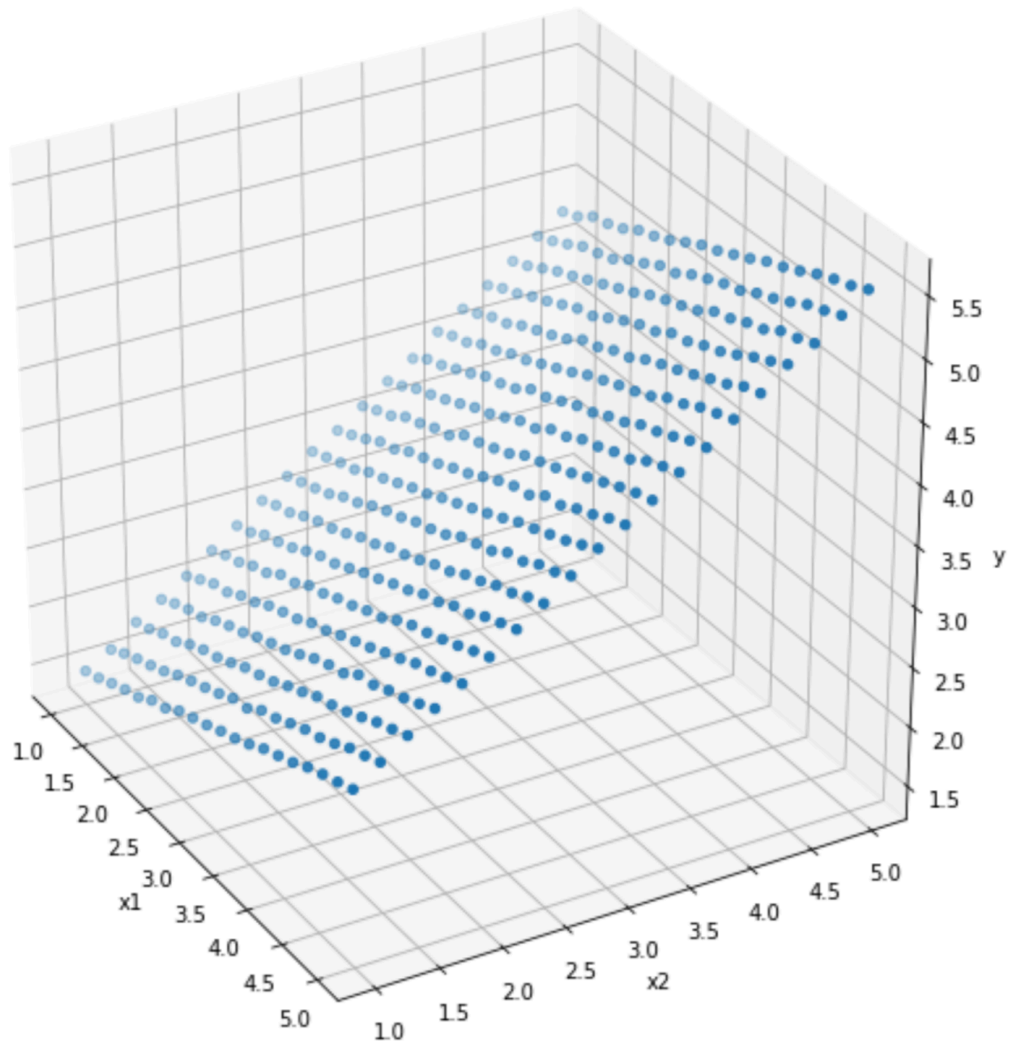
↓ 원핫 인코딩으로 변형

$$y = \sigma(\\ w_{\text{고객번호}_1\text{번}} x_{\text{고객번호}_1\text{번}} + w_{\text{고객번호}_2\text{번}} x_{\text{고객번호}_2\text{번}} + \cdots + w_{\text{고객번호}_{973}\text{번}} x_{\text{고객번호}_{973}\text{번}} + \\ w_{0\sim4\text{세}} x_{0\sim4\text{세}} + w_{5\sim9\text{세}} x_{5\sim9\text{세}} + w_{10\sim14\text{세}} x_{10\sim14\text{세}} + \cdots + w_{65\sim69\text{세}} x_{65\sim69\text{세}} + \\ w_{\text{남}} x_{\text{남}} + w_{\text{여}} x_{\text{여}} + \\ w_{\text{영화번호}_1\text{번}} x_{\text{영화번호}_1\text{번}} + w_{\text{영화번호}_2\text{번}} x_{\text{영화번호}_2\text{번}} + \cdots + w_{\text{영화번호}_{1682}\text{번}} x_{\text{영화번호}_{1682}\text{번}} + \\ w_{1922\text{년}} x_{1922\text{년}} + w_{1923\text{년}} x_{1923\text{년}} + \cdots + w_{1998\text{년}} x_{1998\text{년}})$$

Q) 15세 남자인 고객번호 4번이 1923년에 개봉한 영화번호 2번을 선호할 확률은?

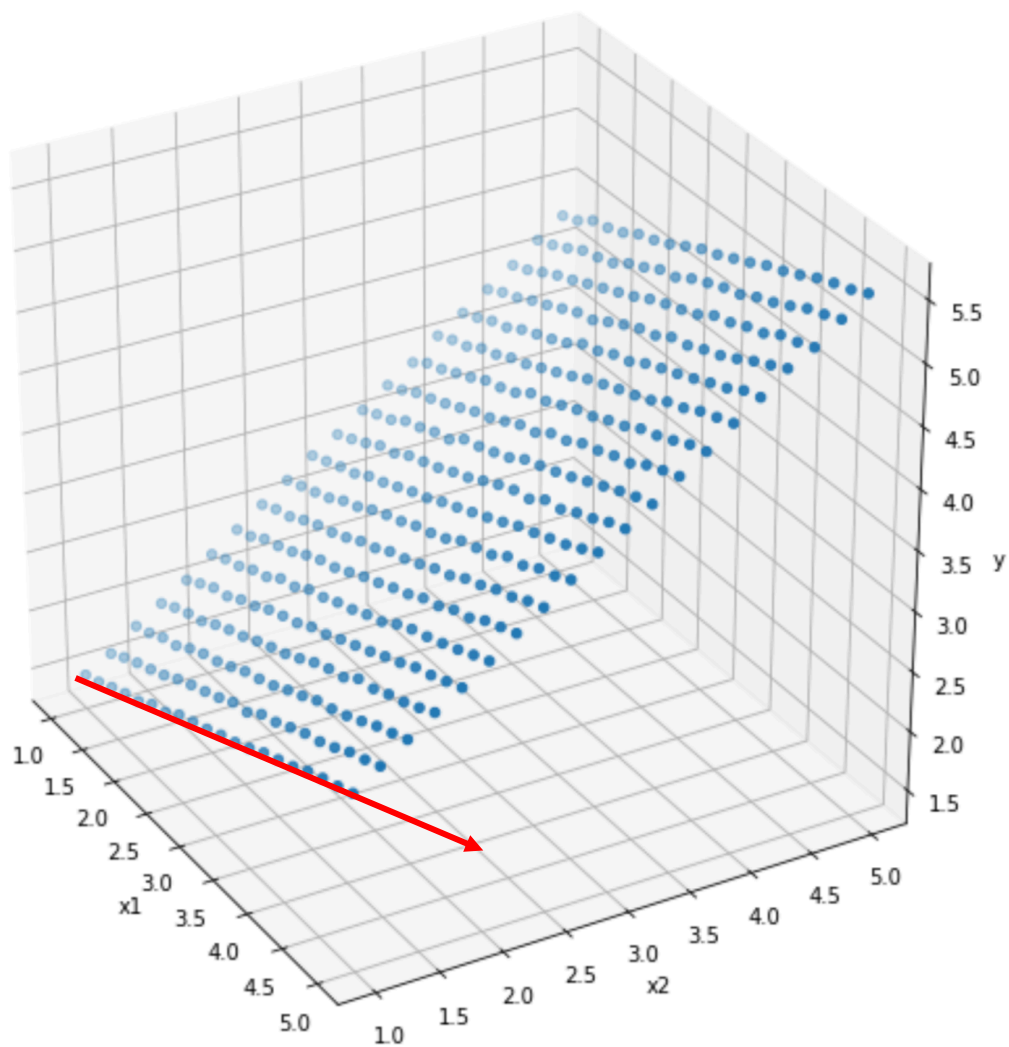
$$y = \sigma(w_{\text{고객번호}_4\text{번}} + w_{15\sim19\text{세}} + w_{\text{남}} + w_{\text{영화번호}_2\text{번}} + w_{1923\text{년}})$$

다변수 선형회귀

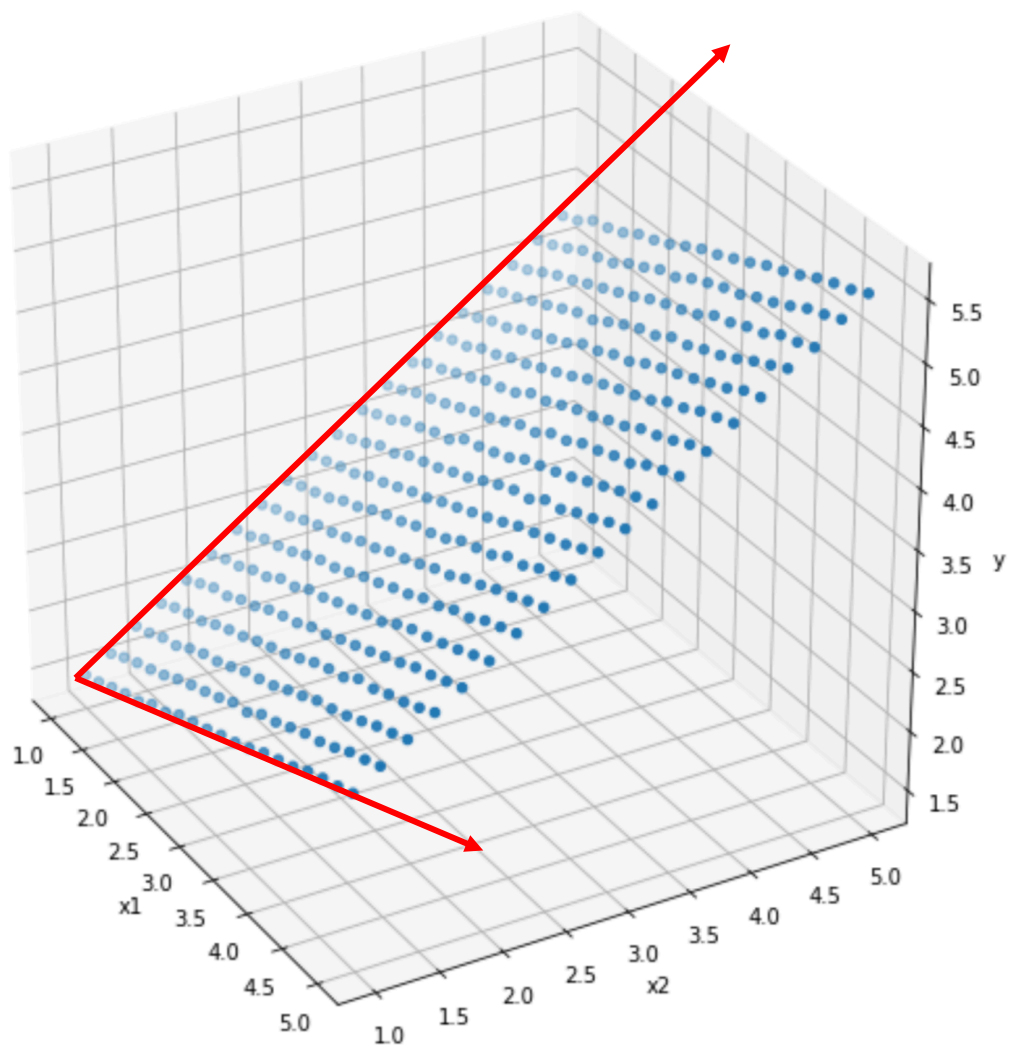


다변수 선형회귀

(1) x_1 이 증가할 때, y 가 직선으로 변환



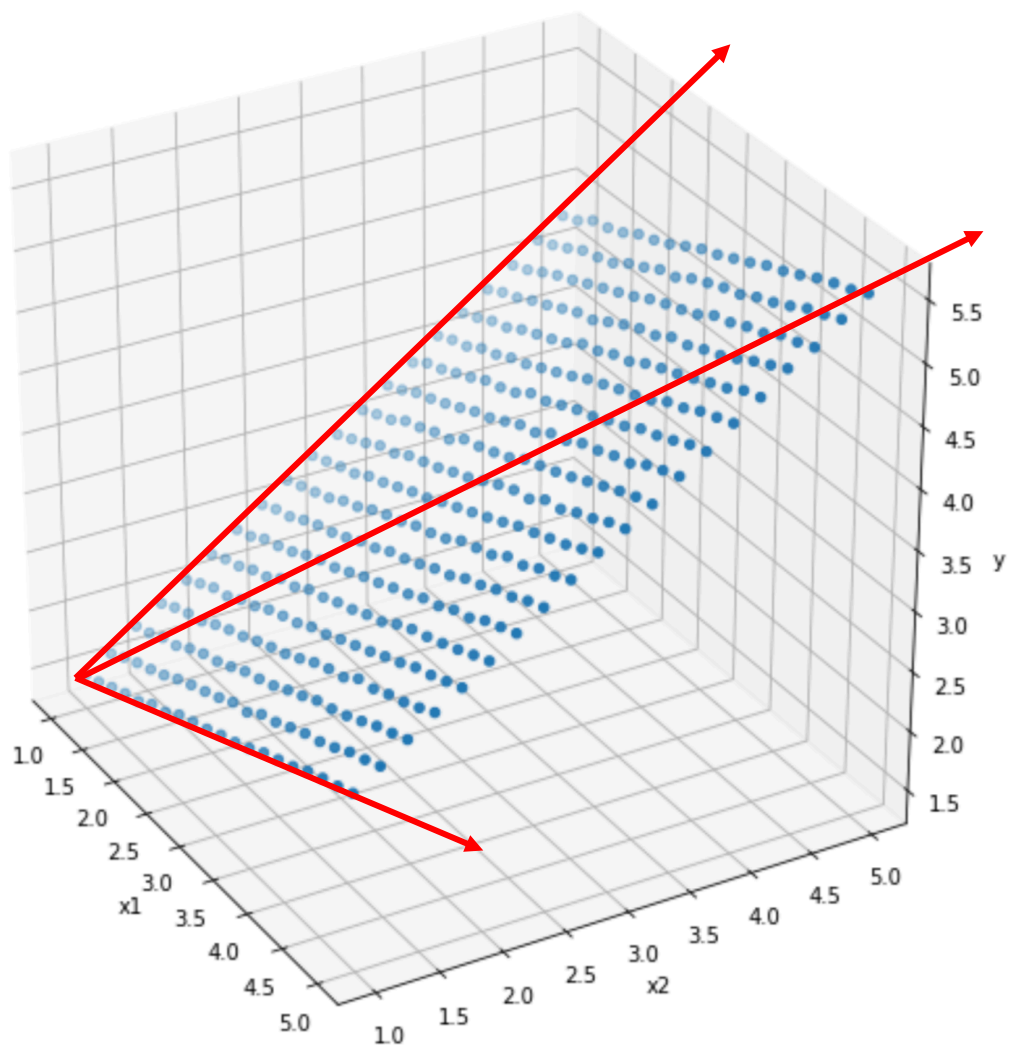
다변수 선형회귀



(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

다변수 선형회귀

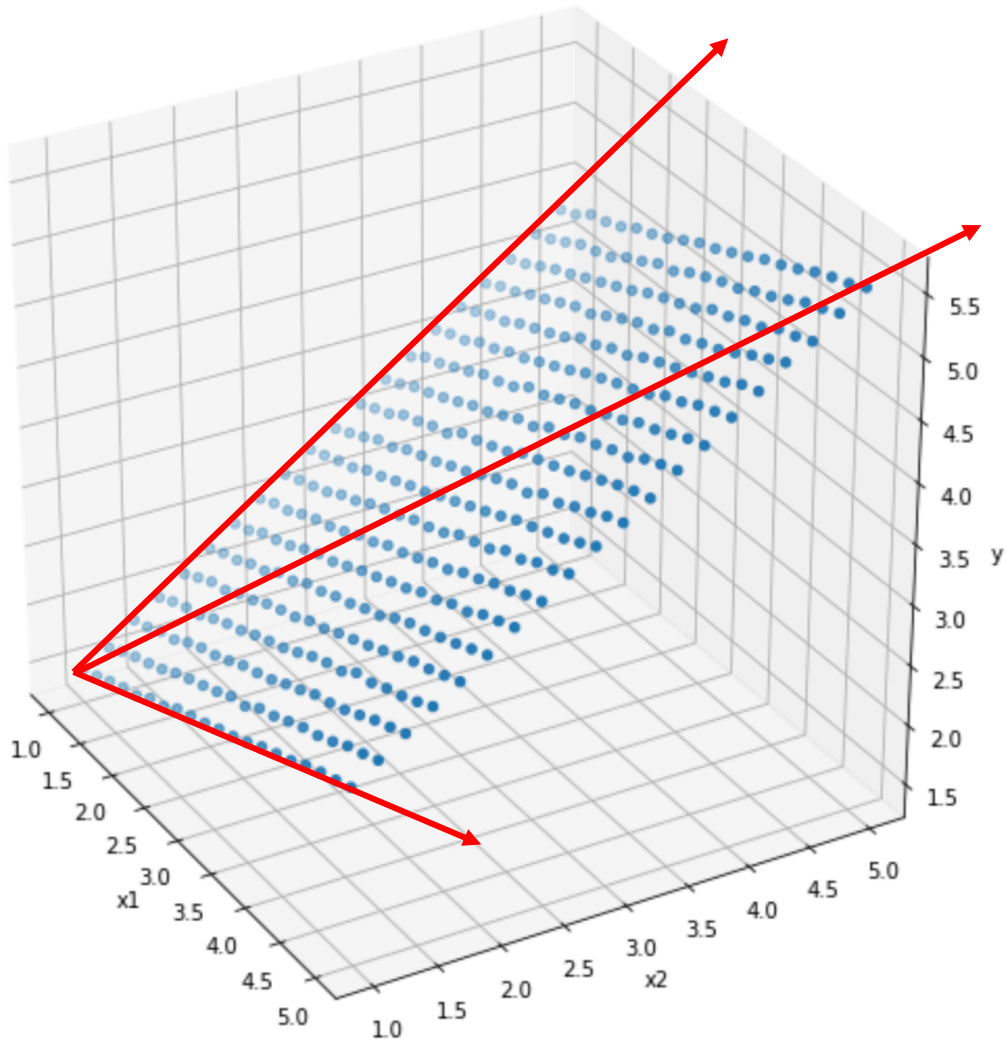


(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

(3) x_1 와 x_2 가 동시에 변할 때, y 가 직선으로 변환

다변수 선형회귀



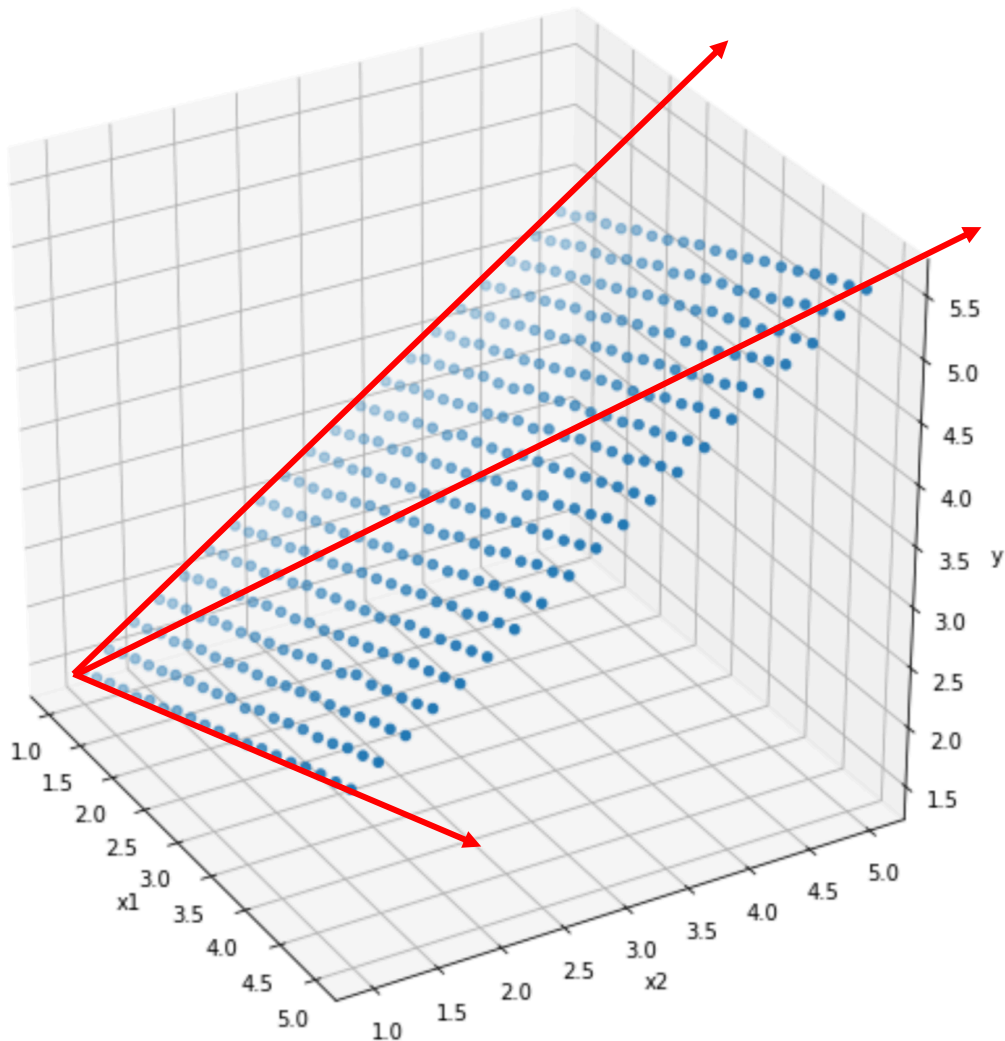
(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

(3) x_1 와 x_2 가 동시에 변할 때, y 가 직선으로 변환

➡ $y = w_1x_1 + w_2x_2 + w_0$

다변수 선형회귀



(1) x_1 이 증가할 때, y 가 직선으로 변환

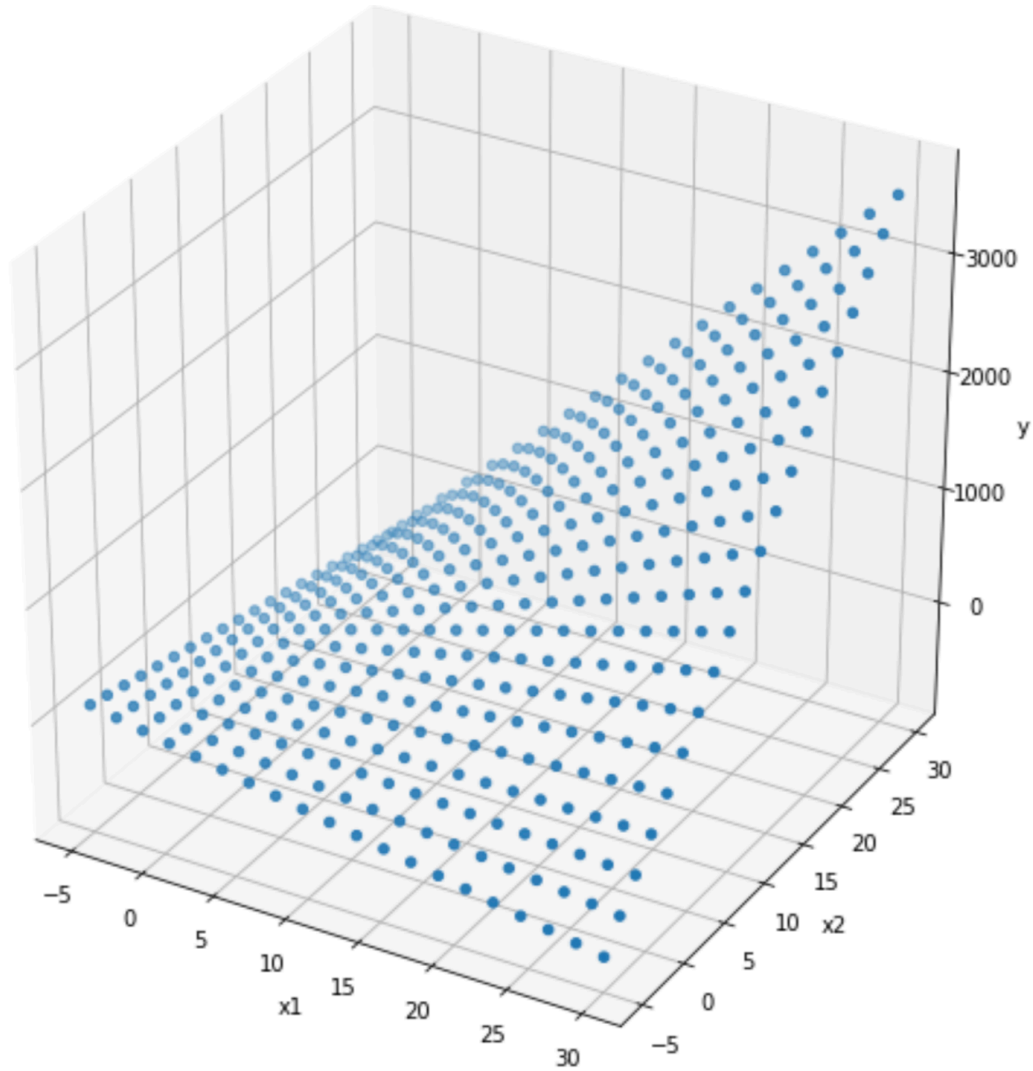
(2) x_2 이 변할 때, y 가 직선으로 변환

(3) x_1 와 x_2 가 동시에 변할 때, y 가 직선으로 변환

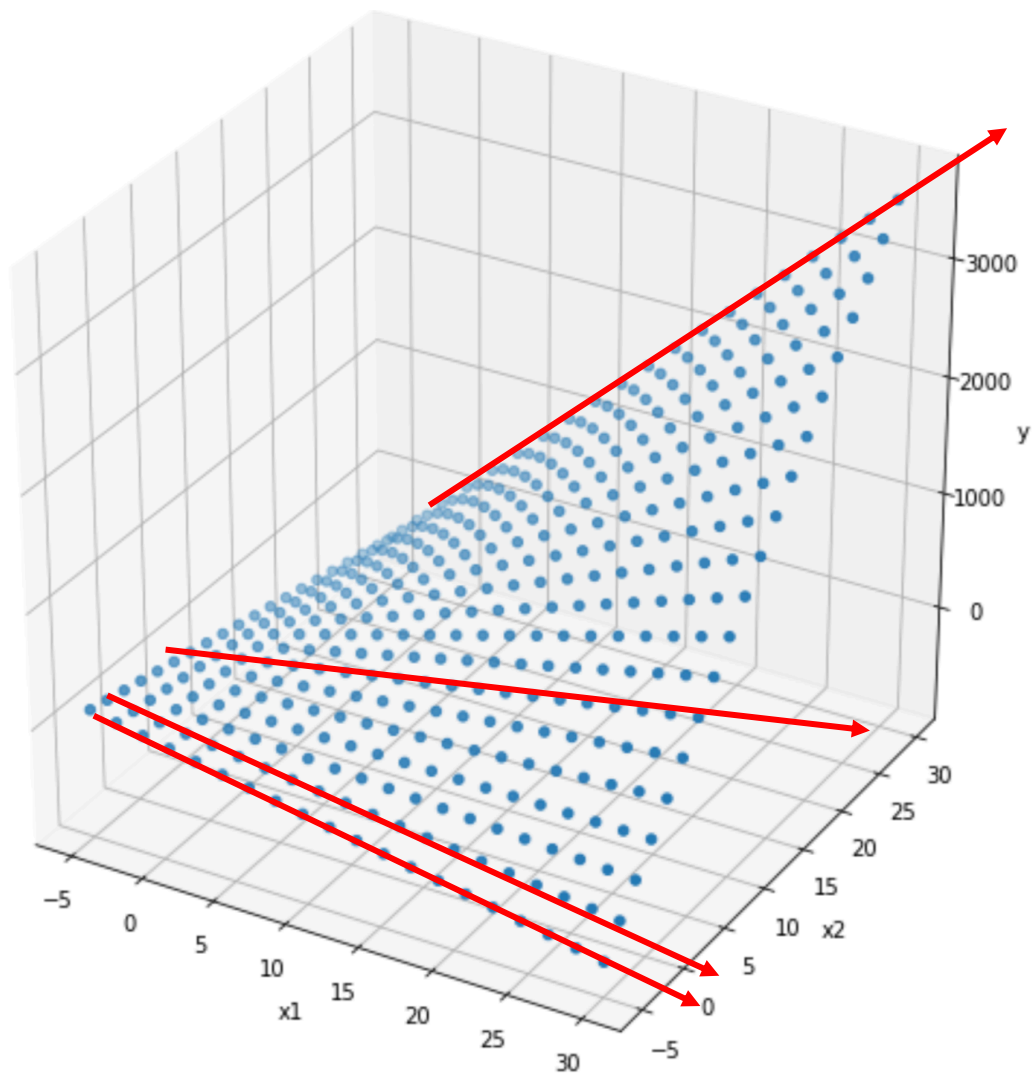
➡ $y = w_1x_1 + w_2x_2 + w_0$

x_1 와 x_2 은 독립적인 관계

다변수 선형회귀 with 상호작용

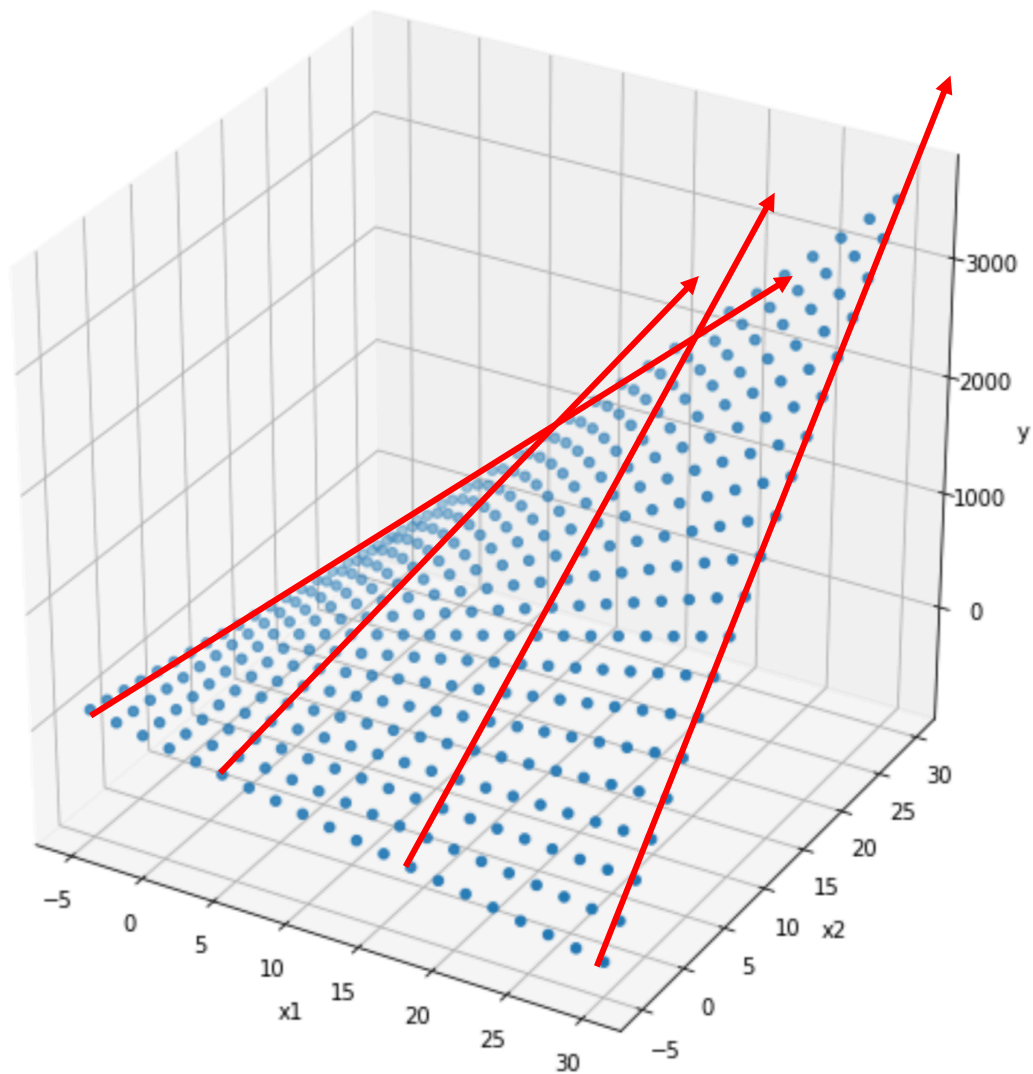


다변수 선형회귀 with 상호작용



(1) x_1 이 증가할 때, y 가 직선으로 변환

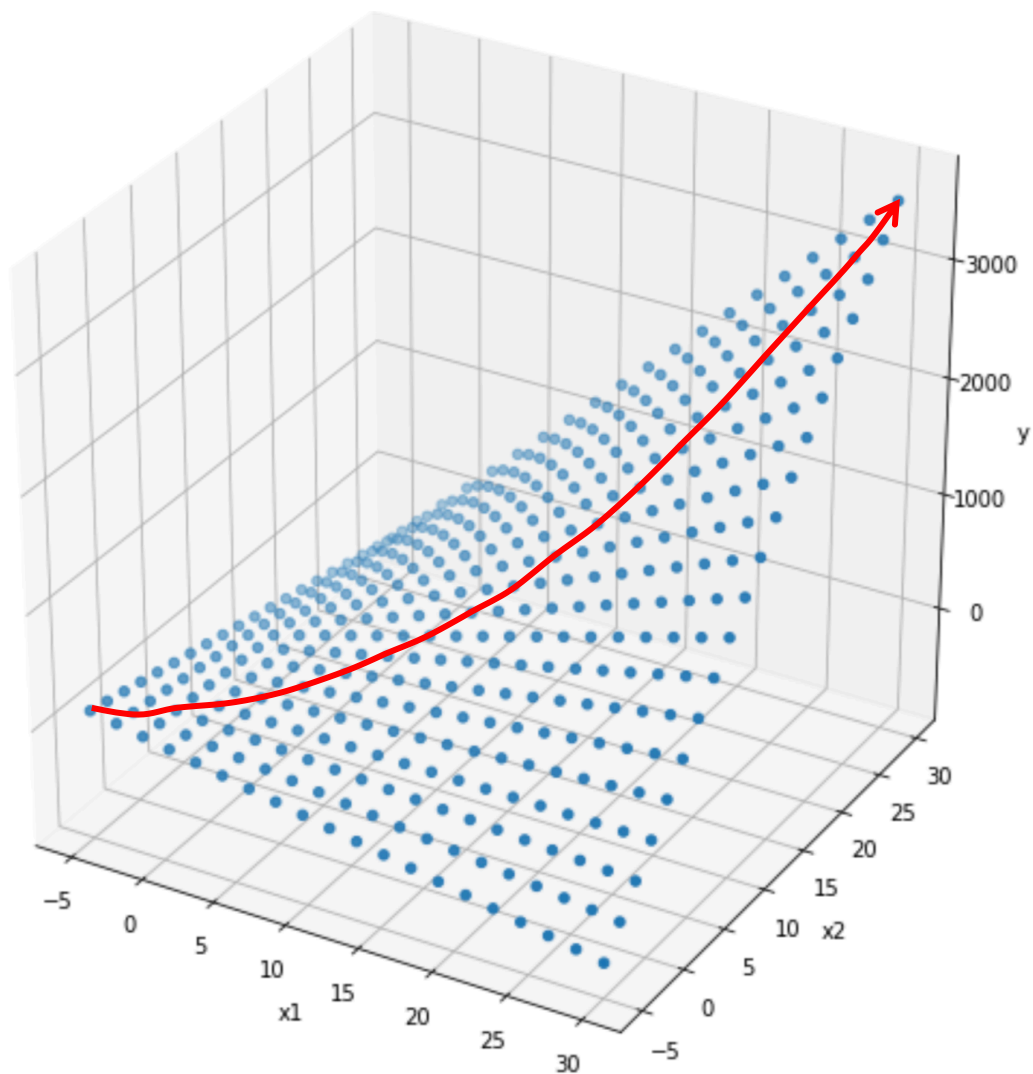
다변수 선형회귀 with 상호작용



(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

다변수 선형회귀 with 상호작용

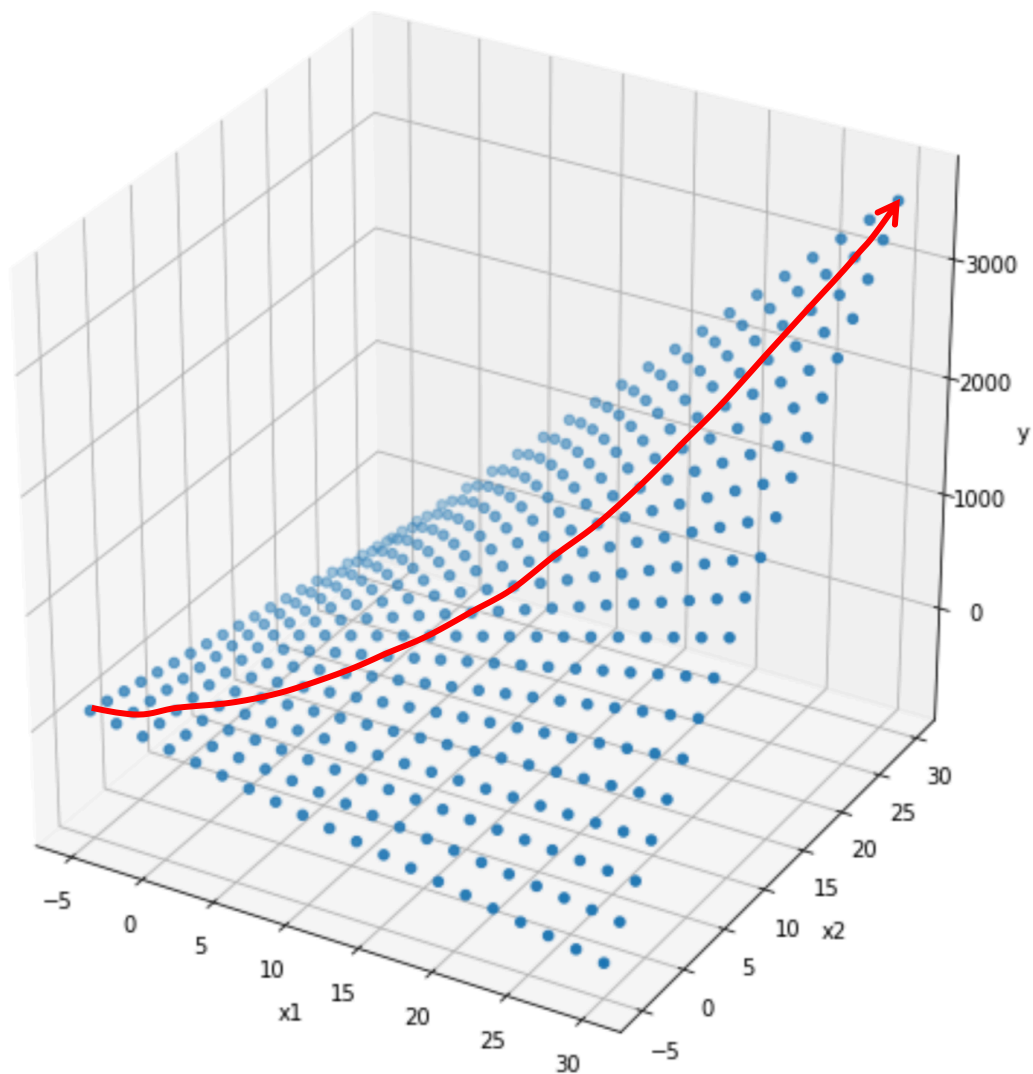


(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

(3) x_1 와 x_2 가 동시에 변할 때, y 가 곡선으로 변환

다변수 선형회귀 with 상호작용



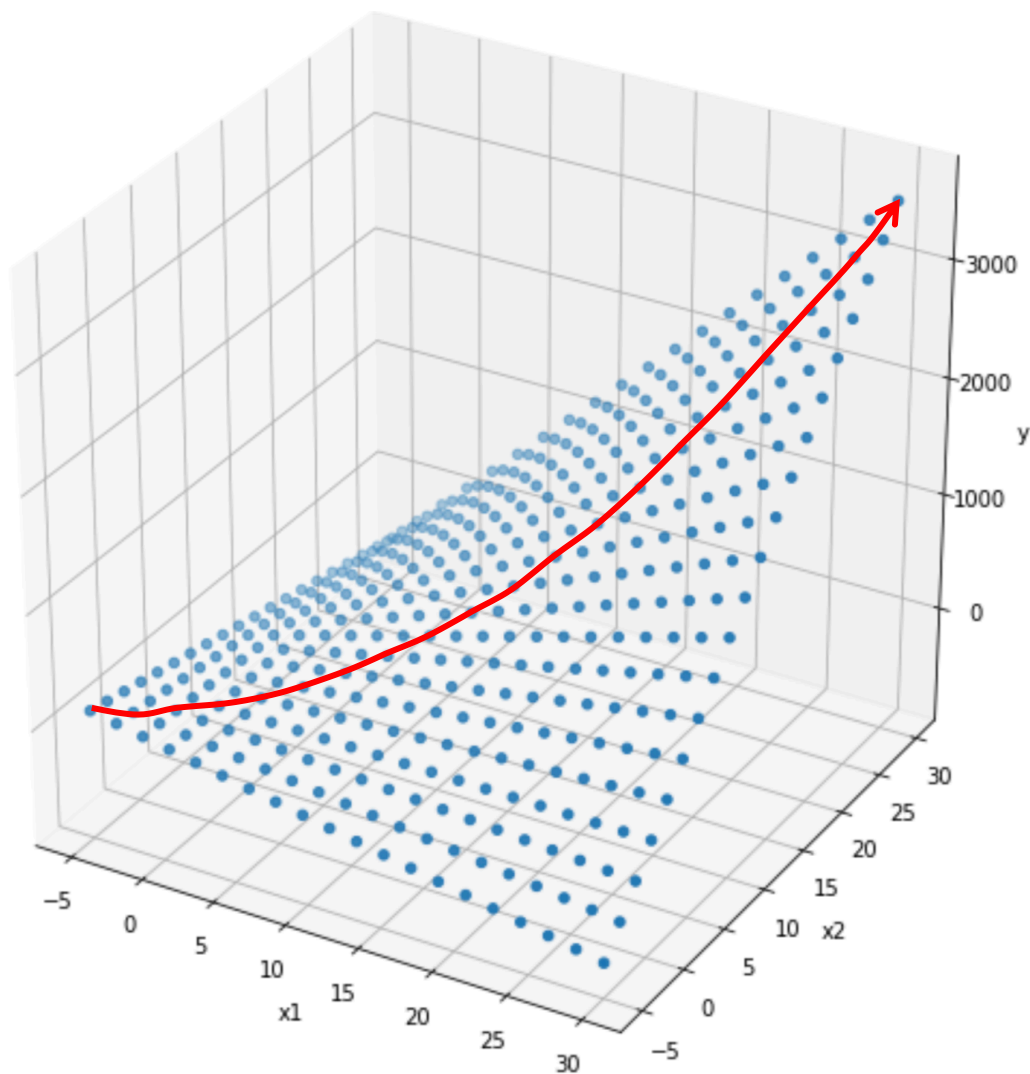
(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

(3) x_1 와 x_2 가 동시에 변할 때, y 가 곡선으로 변환

➡ x_1 와 x_2 사이에는 상호작용이 존재

다변수 선형회귀 with 상호작용



(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

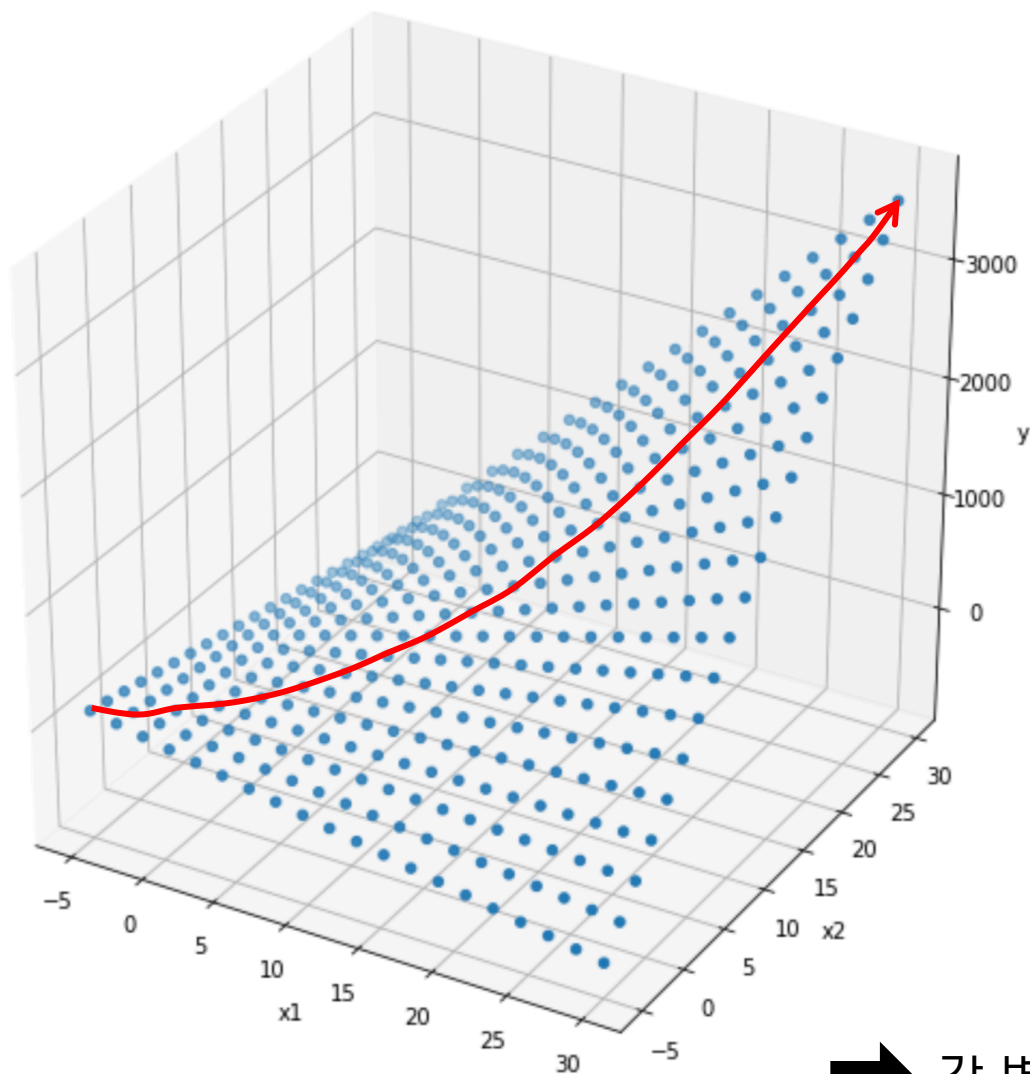
(3) x_1 와 x_2 가 동시에 변할 때, y 가 곡선으로 변환

➡ x_1 와 x_2 사이에는 상호작용이 존재

$$y = w_1x_1 + w_2x_2 + w_3\underline{x_1x_2} + w_0$$

x_1 과 x_2 에 대한 상호작용

다변수 선형회귀 with 상호작용



(1) x_1 이 증가할 때, y 가 직선으로 변환

(2) x_2 이 변할 때, y 가 직선으로 변환

(3) x_1 과 x_2 가 동시에 변할 때, y 가 곡선으로 변환

➡ x_1 과 x_2 사이에는 상호작용이 존재

$$y = w_1x_1 + w_2x_2 + w_3\underline{x_1x_2} + w_0$$

x_1 과 x_2 에 대한 상호작용

➡ 각 변수 간 상호작용도 학습할 수 있는 모델이 필요

명목형 변수를 위한 선형회귀 with 상호작용

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_{\text{장르}} X_{\text{장르}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 학습하기 위한 가중치도 필요

명목형 변수를 위한 선형회귀 with 상호작용

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_{\text{장르}} X_{\text{장르}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 학습하기 위한 가중치도 필요

상호작용 가짓 수

(고객, 나이)

(나이, 성별)

(성별, 연도)

(고객, 성별)

(나이, 영화)

(성별, 장르)

(고객, 영화)

(나이, 연도)

(영화, 연도)

(고객, 연도)

(나이, 장르)

(영화, 장르)

(고객, 장르)

(성별, 영화)

(연도, 장르)

명목형 변수를 위한 선형회귀 with 상호작용

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_{\text{장르}} X_{\text{장르}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 학습하기 위한 가중치도 필요

상호작용 가짓 수

(고객, 나이) = $943 * 12$	(나이, 성별) = $12 * 2$	(성별, 연도) = $2 * 77$
(고객, 성별) = $943 * 2$	(나이, 영화) = $12 * 1682$	(성별, 장르) = $2 * 20$
(고객, 영화) = $943 * 1682$	(나이, 연도) = $12 * 77$	(영화, 연도) = $1682 * 77$
(고객, 연도) = $943 * 77$	(나이, 장르) = $12 * 20$	(영화, 장르) = $1682 * 20$
(고객, 장르) = $943 * 20$	(성별, 영화) = $2 * 1682$	(연도, 장르) = $77 * 20$

명목형 변수를 위한 선형회귀 with 상호작용

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_{\text{장르}} X_{\text{장르}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 학습하기 위한 가중치도 필요

상호작용 가짓 수

(고객, 나이) = 943 * 12	(나이, 성별) = 12 * 2	(성별, 연도) = 2 * 77
(고객, 성별) = 943 * 2	(나이, 영화) = 12 * 1682	(성별, 장르) = 2 * 20
(고객, 영화) = 943 * 1682	(나이, 연도) = 12 * 77	(영화, 연도) = 1682 * 77
(고객, 연도) = 943 * 77	(나이, 장르) = 12 * 20	(영화, 장르) = 1682 * 20
(고객, 장르) = 943 * 20	(성별, 영화) = 2 * 1682	(연도, 장르) = 77 * 20

총 상호작용 가짓수 = 1,880,423 개

명목형 변수를 위한 선형회귀 with 상호작용

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_{\text{장르}} X_{\text{장르}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 학습하기 위한 가중치도 필요

상호작용 가짓 수

(고객, 나이) = 943 * 12	(나이, 성별) = 12 * 2	(성별, 연도) = 2 * 77
(고객, 성별) = 943 * 2	(나이, 영화) = 12 * 1682	(성별, 장르) = 2 * 20
(고객, 영화) = 943 * 1682	(나이, 연도) = 12 * 77	(영화, 연도) = 1682 * 77
(고객, 연도) = 943 * 77	(나이, 장르) = 12 * 20	(영화, 장르) = 1682 * 20
(고객, 장르) = 943 * 20	(성별, 영화) = 2 * 1682	(연도, 장르) = 77 * 20

총 상호작용 가짓수 = 1,880,423 개  가중치가 지나치게 많다!

Factorization Machine의 아이디어

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 계산하기 위한 임베딩 행렬(Factor)을 각각 구성하자

상호작용 임베딩 행렬

고객번호					나이					성별					영화번호					연도				
1					0~4					남					1					1922				
2					5~9					녀					2					1923				
3					10~14										3					1924				
...								
943					65~69										1682					1998				

Factorization Machine의 아이디어

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 계산하기 위한 임베딩 행렬(Factor)을 각각 구성하자

상호작용 임베딩 행렬

고객번호					나이					성별					영화번호					연도				
1	a1	a2	a3	a4	0~4					남					1	b1	b2	b3	b4	1922				
2					5~9					녀					2					1923				
3					10~14										3					1924				
...								
943					65~69										1682					1998				

Q) 고객 1번과 영화번호의 상호작용 효과는?

$$\text{고객번호} \circ \text{영화번호} = a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4$$

Factorization Machine의 아이디어

$$y = \sigma(W_{\text{고객번호}} X_{\text{고객번호}} + W_{\text{나이}} X_{\text{나이}} + W_{\text{성별}} X_{\text{성별}} + W_{\text{영화번호}} X_{\text{영화번호}} + W_{\text{연도}} X_{\text{연도}} + W_0)$$

상호작용도 학습시키고 싶다! => 상호작용을 계산하기 위한 임베딩 행렬(Factor)을 각각 구성하자

상호작용 임베딩 행렬

고객번호					나이				성별				영화번호				연도								
1	a1	a2	a3	a4	0~4	c1	c2	c3	c4	남 녀					1					1922					
2					5~9											2					1923				
3					10~14											3					1924				
...								
943					65~69											1682					1998				

Q) 고객 1번과 나이 0~4세의 상호작용 효과는?

$$\text{고객번호} \circ \text{나이} = a_1 c_1 + a_2 c_2 + a_3 c_3 + a_4 c_4$$

Factorization Machine의 아이디어

Factorization Machine의 수식

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

상호작용 임베딩 행렬

고객번호					나이					성별		영화번호					연도				
1					0~4					남		1					1922				
2					5~9					녀		2					1923				
3					10~14							3					1924				
...								
943					65~69							1682					1998				

Factorization Machine의 아이디어

Factorization Machine의 수식

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

상호작용 임베딩 행렬

고객번호				나이				성별				영화번호				연도			
1				0~4				남				1				1922			
2				5~9				녀				2				1923			
3				10~14								3				1924			
...						
943				65~69								1682				1998			

Q) 나이가 13세인 여성 1번 고객이 1922년에 출시한 액션 영화 3번에 대한 상호작용 효과?

$\langle \text{고객}_1, \text{나이}_{10\sim14\text{세}} \rangle + \langle \text{고객}_1, \text{성별}_{\text{여}} \rangle + \langle \text{고객}_1, \text{영화}_3 \rangle + \langle \text{고객}_1, \text{연도}_{1992} \rangle + \dots$
=> 각 임베딩 벡터끼리의 DOT 연산의 합