$$X = \{x_i\}_{i=1}^{\ell} \qquad a: X \to \tilde{X}$$

$$X \in \mathbb{R}^D$$
$$\tilde{X} \in \mathbb{R}^d \qquad d < D$$

1) Выкидывать признаки
   Отбор признаков

2) Генерируем новые признаки основываясь на старых
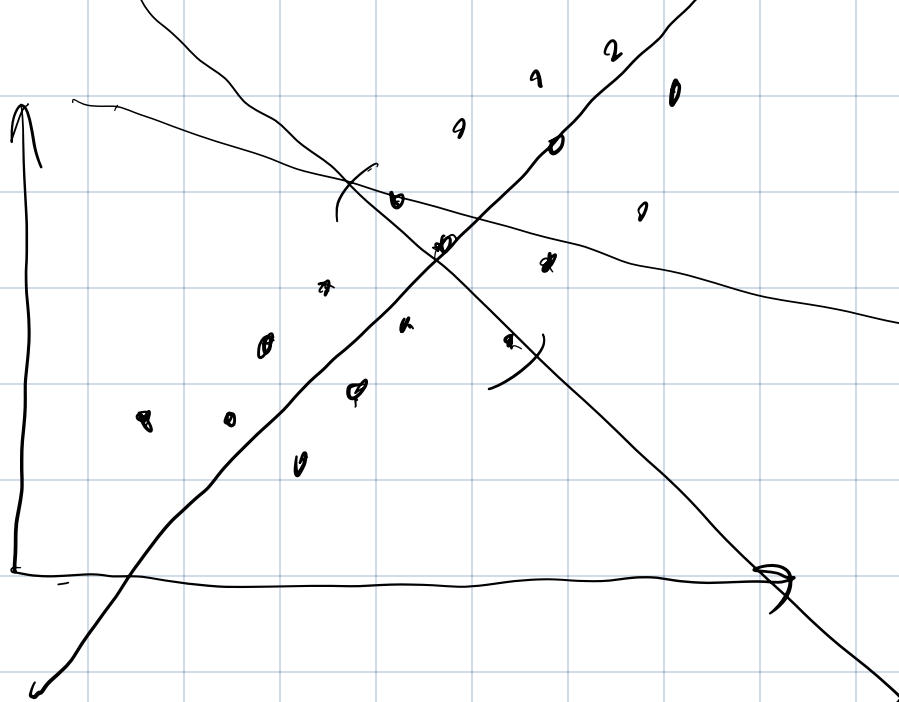
PCA (Метод главных компонент)

$$X \qquad\qquad z_{ij} = \sum_{k=1}^{D} W_{jk} \, x_{ik}$$

новый признак   старый признак

$$- 3$$

$$z_i = W_1 x_{i_1} + W_2 x_{i_2}$$

$$\begin{cases} \sum_{j=1}^{d} w_j^T X^T X w_j \rightarrow \max_w \\ W^T W = I \end{cases}$$

$$X^T X \quad d$$

U - пользователи $\{u_1, \dots, u_n\}$

I - items $\{i_1, \dots, i_m\}$

$(u, i) \in U \times I \quad \exists \; r_{ui}$ - рейтинг.

целевая переменная

1) $r_{ui} \in \{0, 1\}$ - клики

2) $r_{ui} \in \mathbb{R}$ - время просмотра

3) $r_{ui} \in \{1, \dots, k\}$ - оценка.

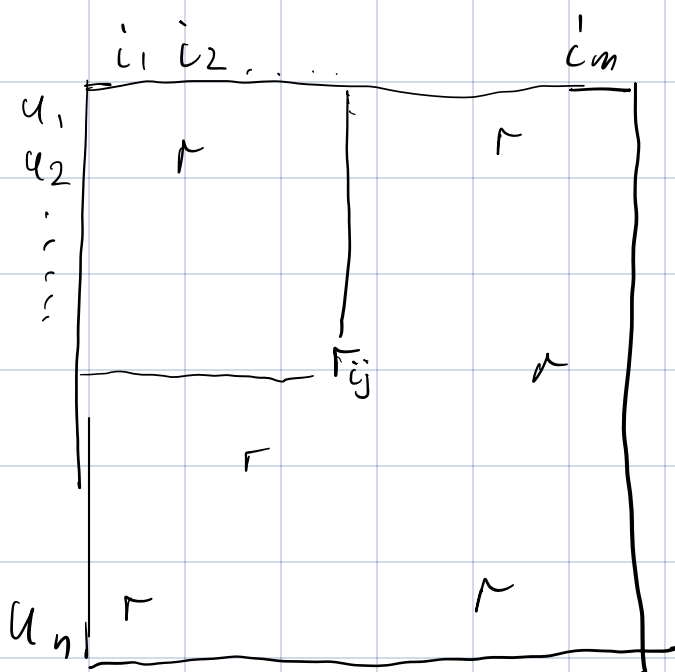$R = \{(u, i) \in U \times I \mid \exists r_{ui}\}$ - обуч.

выборка

$$|R| << |U \times I|$$

Задача: построить $a : U \times I \to Y$

$$a(u, i) = r_{ui} \quad \forall u, i$$

① Коллаборативная фильтрация

Используем <u>только!</u> информацию
о взаимодействии U и I

Memor - based.

① User - based.

$u, \vartheta \in U$

$$I_{u\vartheta} = \{ i \in I \mid \nexists r_{ui} \ \& \ \exists r_{\vartheta i} \}$$

$$Sim(u, \vartheta) = \frac{\sum_{i \in I_{u\vartheta}} (r_{ui} - \bar{r}_u)(r_{\vartheta i} - \bar{r}_\vartheta)}{\sqrt{\sum_{i \in I} (r_{ui} - \bar{r}_u)^2 \sum} (r_\vartheta - \bar{r})}$$

$$\overline{\Gamma_4} \quad \overline{\Gamma_\upsilon}$$

$$u_0 \to \underline{U(u_0)} = \{u \in U \mid sim(u_0, u) > \alpha\}$$

$$\uparrow$$

коллаборация

$$P_i = \frac{|\{u \in U(u_0) \mid \Gamma_{ui}\}|}{|U(u_0)|}$$

рекомендуем для u_0 к товаров
с максим. P_i
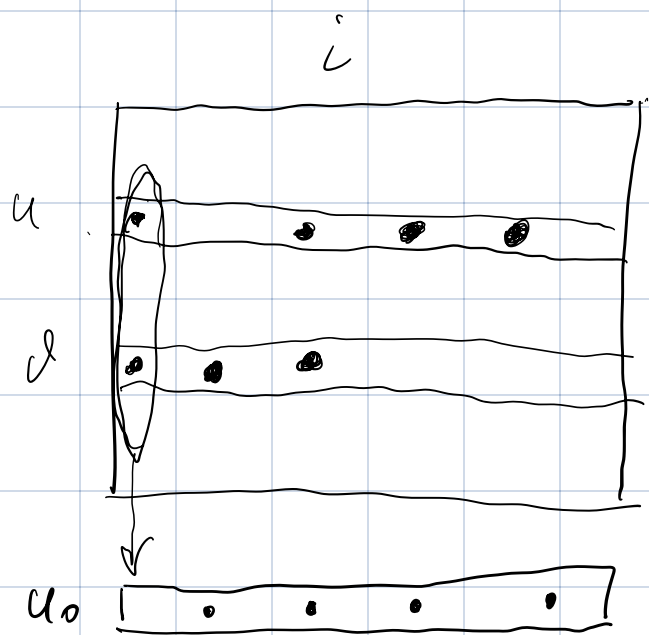
$$sim(u_0, u) > \alpha$$

$$U(u_0).$$

item-based.

$$I(u_0) = \{ i \in I \mid \exists \Gamma_{u_0 i_0}, \; \underline{sim(i_0, i) > \alpha} \}$$

$$P_i = \max_{i_0: \; \exists \Gamma_{u_0 i_0}} sim(i_0, i)$$

рекомендуем пользователю i_0 с максим.
P_i

⊕ простые

— доработать
— матрица User-Item большая.
— нет обучения

② Модели со скрытыми переменными
LFM (latent factor model)

$$U \to P_u \in \mathbb{R}^d$$

$$i \to q_i \in \mathbb{R}$$

$$\langle p_u, q_i \rangle \approx r_{ui}$$

$$\sum_{(u,i) \in R} L\left(r_{ui}, \langle p_u, q_i \rangle\right) \to \min_{P, Q}$$

$$P = (p_1 | \ldots | p_m)$$

$$Q = (q_1 | \ldots | q_n)$$

$$\sum_{(u,i) \in R} \left(r_{ui} - \langle p_u, q_i \rangle\right)^2 \to \min_{P, Q}$$

Как обучать LFM?

① SGD.

$(u, i) \in R$ — случайное

$$p_u = p_u - 2\left(\langle p_u, q_i \rangle - r_{ui}\right) q_i$$

$$q_i = q_i - \eta \left(\langle p_u, q_i \rangle - r_{ui} \right) p_u.$$

② ALS.

$$p_u = \left(\sum_{i: \exists r_{ui}} q_i q_i^T \right)^{-1} \sum_{i: \exists r_{ui}} r_{ui} q_i$$

$$q_i = \left(\sum_{u: \exists r_{ui}} p_u p_u^T \right)^{-1} \sum_{u: \exists r_{ui}} r_{ui} p_u$$