OTUS

ОНЛАЙН-ОБРАЗОВАНИЕ

Матричные Разложения

Дмитрий Музалевский
Преподаватель

План на сегодня

1. **Матричные разложения. Ранг матрицы. Приближение**
2. SGD и ALS. Прогнозирование значений. Implicit методы.
3. Практика

Спектральное разложение

$$X = S^T \cdot D \cdot S,$$

где S — ортогональная матрица, а $D = \mathrm{diag}(\lambda_1, ..., \lambda_i)$ — диагональная матрица из собственных значений матрицы X.

$$f(y) = y^T \cdot S^T \cdot D \cdot S \cdot y = (S \cdot y)^T \cdot D \cdot (S \cdot y) = z^T \cdot D \cdot z = \sum_{i=1}^{n} \lambda_i z_i^2,$$

где была введена естественная замена $z = S \cdot y$.

Сингулярное разложение

В случае произвольной матрицы X имеет место так называемое сингулярное разложение. Пусть X — произвольная матрица, тогда существуют такие ортогональные матрицы U и V, а также диагональная матрица D, что:

$$X = U \cdot D \cdot V.$$

Ранг Матрицы

Рангом (строчным и столбцовым) матрицы называется соответственно максимальное количество линейно независимых строк или столбцов. Одним из ключевых результатов линейной алгебры является то, что строчный ранг совпадает со столбцовым рангом и равен максимальному размеру невырожденной подматрицы. То есть ранг матрицы $X \in \mathbb{R}^{n \times m}$ размера $n \times m$ не может превосходить ни число строк, ни число столбцов в этой матрице:

$$X \in \mathbb{R}^{n \times m} \implies \operatorname{rg}(X) \leq \min(n, m).$$

Ранг Матрицы

$$X = AB, \qquad A \in \mathbb{R}^{n\times k}, \qquad B \in \mathbb{R}^{k\times m}.$$

В практических задачах данные всегда измеряются вместе с шумом и поэтому не вся информация, содержащаяся в матрице с данными представляет ценность для исследователя. Таким образом, ставится задача о нахождении лучшей аппроксимации исходной матрицы X некоторой матрицей, ранг которой не превосходит $r < k$. Любая матрица ранга r может быть представлена как произведение следующих матриц:

$$U \cdot V^T, \qquad U \in \mathbb{R}^{m\times k}, \quad V \in \mathbb{R}^{n\times k}.$$

Норма Матрицы

$$\|A\|_2 = \sqrt{\sum_{i,j} |a_{ij}|^2}.$$

В конечном итоге, задача приближения матрицы матрицей меньшего ранга примет вид:

$$U, V = \underset{U \in \mathbb{R}^{m \times k}, V \in \mathbb{R}^{n \times k}}{\operatorname{argmin}} \sum_{i,j} \left(x_{ij} - u_i v_j^T\right)^2,$$

а искомой матрицей, дающей наилучшее приближение при заданном ранге, будет матрица UV^T.

Преобразование признаков

Пусть X — матрица признаков объектов. Пусть для X построено наилучшее приближение матрицей UV^T ранга $k < n$, где n — количество признаков в исходной задаче.

Матрица U может быть проинтерпретирована как матрица новых признаков тех же объектов. При этом размерность пространства признаков уменьшается — происходит сжатие с потерями, причем теряется минимум полезной информации.

Рекомендации

Пусть X — матрица с оценками, которые поставил или поставил бы пользователь под номером i фильму под номером j. Поскольку далеко не все пользователи смотрели все фильмы и выставили оценку, известны не все элементы этой матрицы.

Чтобы спрогнозировать неизвестные данные, можно попытаться приблизить исходную матрицу с помощью матрицы меньшего ранга. В таком случае в качестве нормы следует использовать норму Фробениуса, где суммирование идет только по известным элементам матрицы X. После того, как наилучшее приближение по известным данным было найдено, эту матрицу можно использовать, чтобы спрогнозировать еще не известные данные.

	Пила	Улица Вязов	Ванильное небо	1 + 1
Маша	5	4	1	2
Юля	5	5	2	
Вова			3	5
Коля	3	?	4	5
Петя				4
Ваня		5	3	3

Сингулярное разложение и низкоранговое приближение

Пусть задана матрица X. Требуется найти такую матрицу, ранг которой $\operatorname{rg} \hat{X} \leq k$, которая наилучшим образом приближает исходную:

$$\hat{X} = \underset{\operatorname{rg} \hat{X} \leq k}{\operatorname{argmin}} \left\| X - \hat{X} \right\|$$

SVD для исходной матрицы имеет вид:

$$X = U \cdot D \cdot V^T,$$

где U и V — ортогональные (то есть и невырожденные) матрицы, а D — диагональная матрица ранга $\operatorname{rg} X$. Можно сделать естественную замену искомой матрицы $\hat{X}$ на матрицу $\hat{D}$ (не обязательно диагональную, поэтому это тождественное преобразование):

$$\hat{X} = U \cdot \hat{D} \cdot V^T.$$

Задача перепишется в виде:

$$\hat{X} = \underset{\operatorname{rg} \hat{D} \leq k}{\operatorname{argmin}} \left\| U \cdot (D - \hat{D}) \cdot V^T \right\| = \underset{\operatorname{rg} \hat{D} \leq k}{\operatorname{argmin}} \left\| D - \hat{D} \right\|.$$

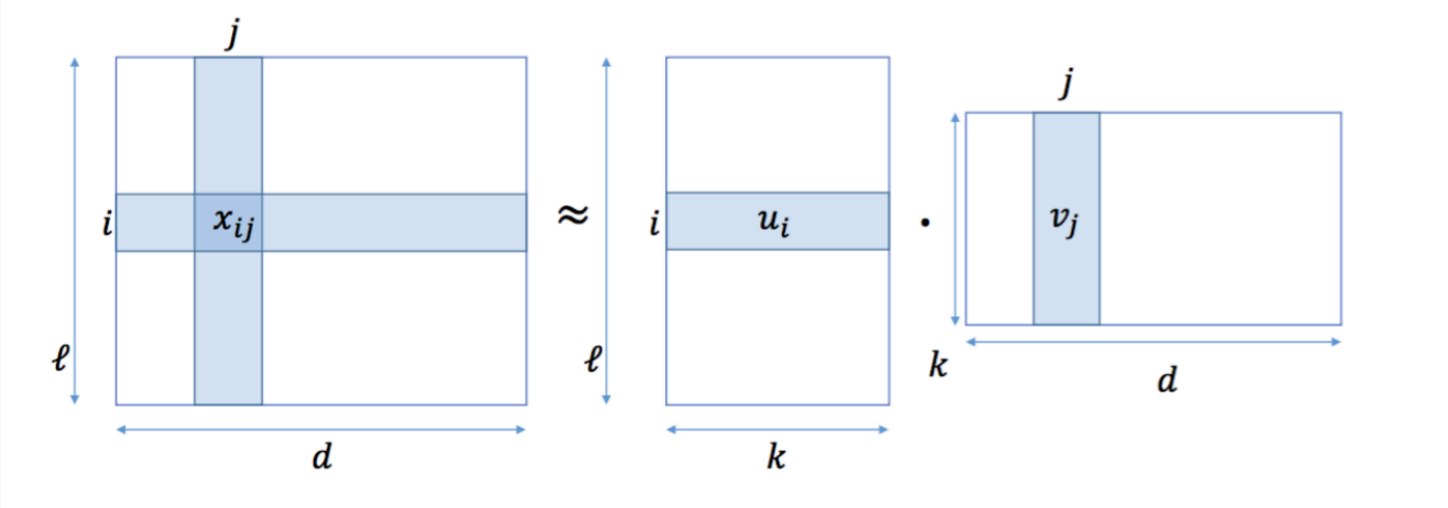
План на сегодня

1. Матричные разложения. Ранг матрицы. Приближение
2. **SGD и ALS. Прогнозирование значений. Implicit методы.**
3. Практика

Матричное разложение: постановка задачи

$$X_{l,n} \approx U_{l,k} \cdot V^T_{k,n},$$

где матрицы U и V задают новое признаковое описание объектов и изначальных признаков.

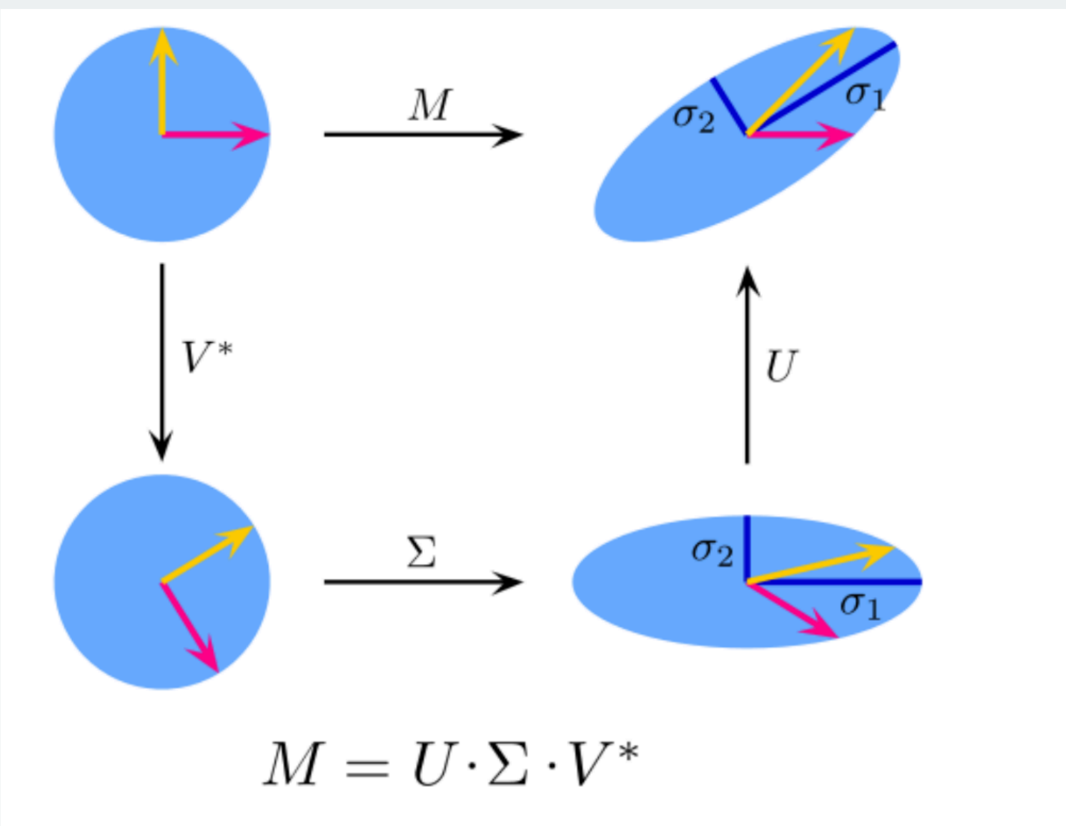


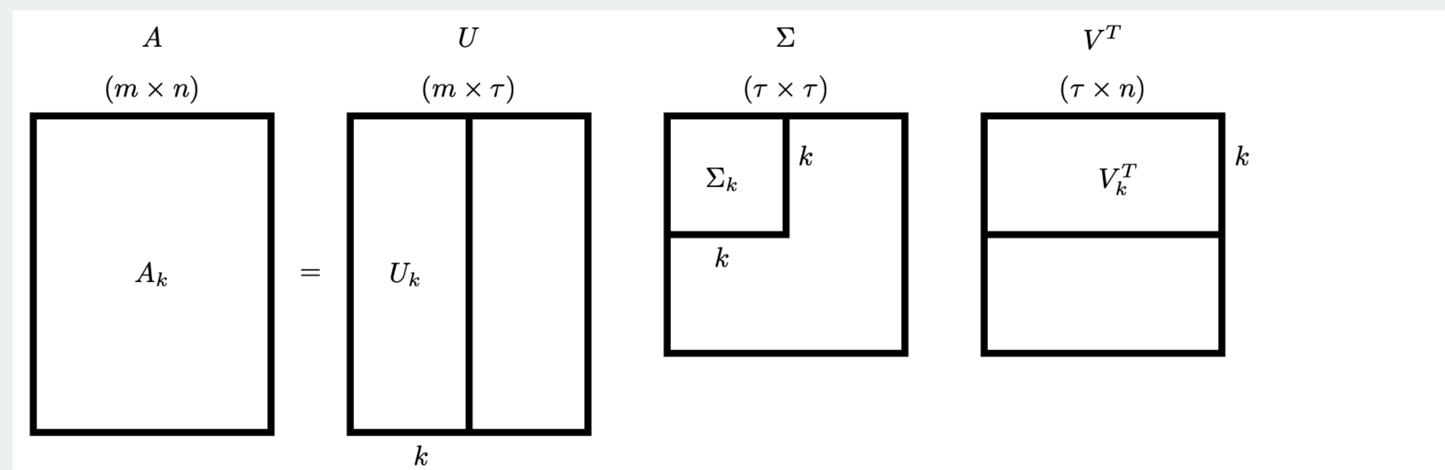
$$\|X - U \cdot V^T\| \to \min.$$

$$\sum_{i,j} (x_{ij} - \langle u_i, v_j \rangle)^2 \to \min,$$

Сингулярное разложение

$$X = U\Sigma V^T,$$





SGD

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 \to \min_{u_i, v_j}.$$

$$\frac{\partial Q}{\partial u_i} = \sum_{i,j} \frac{\partial}{\partial u_i} (\langle u_i, v_j \rangle - x_{i,j})^2 = \sum_j 2(\langle u_i, v_j \rangle - x_{ij}) \frac{\partial \langle u_i, v_j \rangle}{\partial u_i} == \sum_j 2(\langle u_i, v_j \rangle - x_{ij}) v_j.$$

$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \sum_j \varepsilon_{ij} v_j.$$

$$v_j^{(t+1)} = v_j^{(t)} - \eta_t \sum_i \varepsilon_{ij} u_i$$

$$u_i^{(t+1)} = u_i^{(t)} - \gamma_t \varepsilon_{ij} v_j,$$

$$v_j^{(t+1)} = v_j^{(t)} - \eta_t \varepsilon_{ij} u_i.$$

ALS и Регуляризация

$$\frac{\partial Q}{\partial u_i} = 0.$$

На следующей итерации находится новое значение v_j:

$$\frac{\partial Q}{\partial v_i} = 0.$$

$$\frac{\partial Q}{\partial u_i} = \sum_j 2(\langle u_i, v_j \rangle - x_{i,j})v_j = 0$$

$$\sum_j v_j \langle v_j, u_i \rangle = \sum_j x_{ij} v_j$$

$$\left(\sum_j v_j v_j^T\right) u_i = \sum_j x_{ij} v_j$$

$$\left(\sum_i u_i u_i^T\right) v_j = \sum_i x_{ij} u_i$$

$$Q = \sum_{i,j} (\langle u_i, v_j \rangle - x_{ij})^2 + \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\|^2 \to \min_{u_i, v_j},$$

Прогнозирование значений

$$x_{ij} \approx \langle u_i, v_j \rangle.$$

$$\sum_{i,j:x_{ij}\neq 0} (\langle u_i, v_j \rangle - x_{ij})^2 \to \min.$$

$$x_{ij} \approx \mu + \langle u_i, v_j \rangle,$$

$$x_{ij} \approx \mu + b_i^u + b_j^v + \langle u_i, v_j \rangle$$

$$\sum_{i,j} (\mu + b_i^u + b_j^v + \langle u_i, v_j \rangle - x_{ij})^2 \to \min$$

$$\sum_{i,j} (\mu + b_i^u + b_j^v + \langle u_i, v_j \rangle - x_{ij})^2 + \alpha \sum_i \|u_i\|^2 + \beta \sum_j \|v_j\| + \gamma \sum_i b_i^{u^2} + \delta \sum_j b_j^{v^2} \to \min.$$

Implicit методы

	Вечернее платье	Поднос для писем	iPhone 6s	Шуба D&G
Маша	1		1	
Юля	1	1		1
Вова		1	1	
Коля	1	?	1	
Петя		1	1	
Ваня			1	1

Implicit методы

$$x_{ij} = 1 \approx \langle u_i, v_j \rangle$$

$$\sum_{i,j:x_{ij}\neq 0} (\langle u_i, v_j \rangle - x_{ij})^2 \to \min$$

$$u_i = \frac{1}{\sqrt{d}}(1 \dots 1)$$

$$v_j = \frac{1}{\sqrt{d}}(1 \dots 1)$$

- Explicit feedback: можно получить как положительные, так и отрицательные примеры (низкие и высокие оценки фильмов, лайки и дизлайки и т.д.)
- Implicit feedback: есть только положительные (лайки, покупки, просмотры) или отрицательные (дизлайки) примеры.

Implicit методы

$$\sum_{i,j} w_{ij}(\langle u_i, v_j\rangle - x_{ij})^2 \to \min$$

$$w_{ij} = 1 + \alpha|x_{ij}|,$$

План на сегодня

1. RNN
2. LSTM
3. **Практика**

Спасибо
за внимание!