

# Сингулярное разложение матриц

## 1. Матричные разложения

### Спектральное разложение

Представление матрицы в виде произведения некоторых других, обладающих определенными свойствами, называется матричным разложением. Примером матричного разложения может служить спектральное разложение симметричной матрицы  $X$ . По теореме о приведении самосопряженных операторов к диагональному виду:

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

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

Часто в приложениях встречаются так называемые квадратичные формы, то есть функции вида  $f(y) = y^T X y$ , где  $X$  — симметричная матрица. С помощью спектрального разложения матрицы  $X$  можно привести квадратичную форму к более простому виду:

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

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:

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

Сингулярное разложение раскрывает геометрическую структуру линейного преобразования, задаваемого матрицей  $X$ : представляет его в виде последовательных вращения, растяжения по осям и еще одного вращения.

Сингулярное разложение имеет множество практических приложений. В том числе и в области анализа данных.

## 2. Приближение матриц меньшего ранга

### Оценка ранга произведения матриц

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:

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

Пусть теперь матрица  $X \in \mathbb{R}^{n \times m}$  представляет собой произведение матриц  $A \in \mathbb{R}^{n \times k}$  и  $B \in \mathbb{R}^{k \times m}$ , причем  $k < \min(n, m)$ . В таком случае  $\text{rg}(A) \leq k$ ,  $\text{rg}(B) \leq k$ , а следовательно и для ранга матрицы  $X = AB$  будет верна оценка  $\text{rg}(X) \leq k$ .

### Ранг матрицы и сжатие без потерь

Более того, всякую матрицу  $X \in \mathbb{R}^{n \times m}$  ранга  $k$  можно представить в виде произведения матриц  $A$  и  $B$  размеров  $n \times k$  и  $k \times m$  соответственно.

**Доказательство:** Поскольку  $\text{rg} X = k$ , в матрице  $X$  существует система из  $k$  линейно независимых столбцов и всякий столбец матрицы  $X$  представим в виде линейной комбинации столбцов этой системы.

Пусть теперь матрица  $B$  составлена из коэффициентов разложения произвольного столбца матрицы  $X$ , а матрица  $A$  — из вышеупомянутой системы столбцов. Тогда по построению:

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

Это и есть искомое матричное разложение. И поскольку по матрицам  $A$  и  $B$  исходная матрица  $X$  восстанавливается точно, говорят, что происходит сжатие без потерь. Именно поэтому о ранге можно говорить как о мере информационной наполненности матрицы.

## Аппроксимация матрицей меньшего ранга

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:

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

Поэтому удобно переформулировать задачу: требуется найти такие матрицы  $U \in \mathbb{R}^{m \times k}$  и  $V \in \mathbb{R}^{n \times k}$ , что матрица  $X$  и  $UV^T$  будут отличаться не сильно. Для оценки степени близости объектов используется понятие нормы. Для оценки близости матриц обычно используется так называется норма Фробениуса:

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

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

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

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.

## Задача 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.

## 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:

$$\hat{X} = \operatorname{argmin}_{\operatorname{rg} \hat{X} \leq k} \|X - \hat{X}\|$$

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  $\text{rg } X$ . 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 переписывается в виде:

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

Поскольку матрица  $D$  — диагональная, все недиагональные элементы в матрице  $\hat{D}$  должны быть равными нулю (в ином случае это может только увеличить норму разницы). В матрице  $\hat{D}$  может быть максимум  $k$  ненулевых элементов (по условию на ранг  $\hat{X}$ ). Поэтому чтобы обеспечить минимум  $\left\| D - \hat{D} \right\|$  выберем  $\hat{D}$  равной матрице  $D$ , в которой все кроме  $k$  наибольших по модулю диагональных элементов заменены нулями.

Таким образом, наилучшим приближением матрицы  $X$  матрицей ранга  $k$  будет

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

где матрица  $\hat{D}$  — это матрица  $D$ , в которой все кроме  $k$  наибольших по модулю диагональных элементов заменены нулями.

Следует отметить, что и матричное разложение  $X = AB$  определено неоднозначно. В первом случае для любой невырожденной матрицы  $R$  нужного размера можно записать:

$$X = AB = AIB = AR^{-1}RB = A'B'.$$

$A'$  и  $B'$  тоже будут образовывать некоторое другое разложение матрицы  $X$ . Это создает целый ряд проблем при решении задач рекомендаций и подробно будет обсуждаться в соответствующем курсе.