Математические модели обработки сигналов

Тема 2: SVD

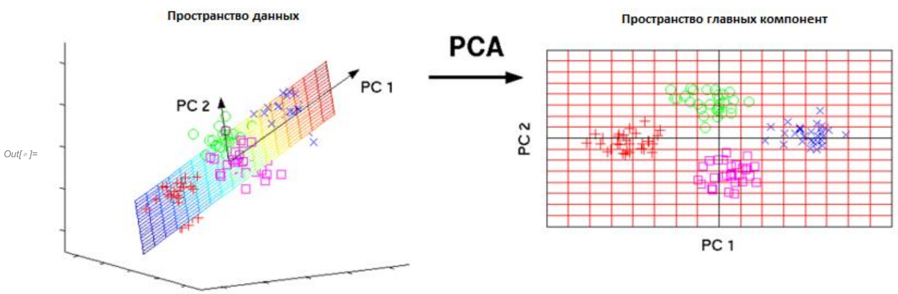
Лектор: Кривошеин А.В.

Мотивация

Ряд тем будет связан с представлением сигналов в различных базисах. Как правило, при работе с сигналами, полученными из одного источника или из нескольких схожих источников, оказывается, что полученные сигналы в некотором смысле “похожи”.

Это может выражаться в том, что при представлении сигналов в подходящем базисе координаты этих сигналов будут содержать большое количество нулей или значений близких к нулю. Тогда несущественные координаты можно отбросить, что фактически позволяет представить исходные сигналы с помощью векторов сравнительно малой размерности.

Такого типа обработка сигналов называется **снижением размерности данных** и может использоваться для сжатия сигналов. Снижение размерности также может быть полезно как первый шаг иных алгоритмов обработки сигналов в задачах анализа сигналов, в задачах классификации, кластеризации и пр.



Алгоритм, который позволяет строить подходящий, лучший в некотором смысле базис для представления сигналов по имеющемуся набору сигналов, является алгоритм, основанный на **сингулярном разложении матриц** (англ. Singular Value Decomposition, SVD).

В данном случае матрица может представлять собой совокупность сигналов одной длины, где сигналы записаны в строках матрицы. Либо это может быть матрица данных, где записаны числовые признаки некоторого набора объектов. То есть каждый объект описывается в матрицы числовым вектором-строкой.

Далее, матрицу данных будем обозначать $A \in \mathbb{C}^{m\times n}$.

Спектральное разложение

Напомним понятие спектрального разложения квадратных матриц. Для матрицы $A \in \mathbb{C}^{m\times m}$ матрицу $A^* = \overline{A^T}$ называют эрмитово-сопряжённой.

Спектральным разложением квадратной матрицы A называют представление этой матрицы с помощью её собственных чисел и собственных векторов.

Напомним, что матрицу $A \in \mathbb{C}^{m\times m}$ называют нормальной, если $A\,A^* = A^*\,A$.

Известно, что матрица A нормальна тогда и только тогда, когда имеет место представление

$$A = Q\,\Lambda\,Q^*,$$

где Q унитарная матрица (то есть $Q\,Q^* = I_m$ или $Q^{-1} = Q^*$), а Λ диагональная матрица с собственными числами матрицы A на диагонали. Это представление и является спектральным разложением.

Иными словами, нормальные матрицы диагонализуемы (и только они).

Кроме того, столбцы $q_1, \; \ldots, \; q_m$ матрицы Q образуют ОНБ пространства $\mathbb{C}^m$ из собственных векторов матрицы A. В этом базисе матрица A имеет наиболее простой диагональный вид.

Если матрица A вещественная и симметричная (то есть $A^T = A$), то все собственные числа этой матрицы вещественны и матрица Q ортогональна (то есть $Q\,Q^T = I_m$ или $Q^{-1} = Q^T$).

Сингулярное разложение

Сингулярное разложения матриц (далее SVD) является аналогом спектрального разложения, который применим в более общем случае, а именно, **применим для любой прямоугольной матрицы**. SVD широко используется в задачах анализа данных и задачах снижения размерности данных, например, для реализации метода главных компонент (англ. Principal Component Analysis, PCA).

Пусть A прямоугольная матрица размера $m \times n$, определим её SVD. Рассмотрим квадратную матрицу B размера $m \times m$ вида

$$B = A A^*.$$

Матрицу B называют матрицей Грамма для матрицы A (поскольку матрица B является матрицей Грамма линейного отображения, действие которого задаётся матрицей A).

Ясно, что матрица B является самосопряжённой, то есть $B = B^*$. Значит матрица B нормальна и диагонализуема. При этом матрица B положительно полу-определена. Действительно,

$$\langle B x, x \rangle = \langle A A^* x, x \rangle = \langle A^* x, A^* x \rangle = \|A^* x\|^2 \geq 0 \quad \forall x \in \mathbb{C}^m.$$

То есть все собственные числа матрицы B неотрицательны.

Сингулярное разложение

Запишем спектральное разложение для матрицы B:

$$B = U \Lambda U^*,$$

где Λ диагональная матрица с собственными числами матрицы B на диагонали, а U унитарная матрица, столбцы которой порождают ОНБ из собственных векторов матрицы B. Обозначим собственные числа за λ_j, $j = 1, \ldots, m$, а соответствующие собственные вектора-столбцы за u_j, то есть

$$\Lambda = \operatorname{diag}\{\lambda_1, \ldots, \lambda_m\}, \quad U = (u_1, \ldots, u_m), \quad B u_j = \lambda_j u_j.$$

Эти матрицы имеют размер $m \times m$. Как отмечено выше, $\lambda_j \geq 0$. Будем также считать, что собственные числа расположены по убыванию значений. Положим $\lambda_j = \sigma_j^2$, где

$$\sigma_1 \geq \ldots = > \sigma_r > \sigma_{r+1} =. .. \sigma_m = 0$$

с соглашением, что $\{\sigma_{r+1}, \ldots, \sigma_m\}$ является пустым множеством при $r = m$. Эти значения σ_j называют **сингулярными числами** матрицы A. Таким образом,

$$\Lambda = \operatorname{diag}\{\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0\}.$$

Сингулярное разложение

Обозначим за $S_r := \operatorname{diag}\{\sigma_1, \ldots, \sigma_r\}$, рассмотрим матрицу размера $m \times n$

$$\Sigma = \begin{pmatrix} S_r & \mathbb{O}_{r\times(n-r)} \\ \mathbb{O}_{(m-r)\times r} & \mathbb{O}_{(m-r)\times(n-r)} \end{pmatrix}.$$

При этом $\Lambda = \Sigma\,\Sigma^*$ и верны равенства

$$B = U\,\Lambda\,U^* = (U\,\Sigma)\,(U\,\Sigma)^*, \quad \text{причём } \operatorname{rank}(\Sigma) = \operatorname{rank}(\Lambda) = \operatorname{rank}(B) = r.$$

Теорема (о полном SVD). Пусть A является матрицей размера $m \times n$ с рангом $\operatorname{rank} A = r$. Тогда найдутся унитарные матрицы U размера $m \times m$ и V размера $n \times n$, такие что

$$A = U\,\Sigma\,V^*,$$

где матрица Σ определена выше.

Матрица U в SVD совпадает с матрицей U полученной из спектрального разложения матрицы B, а матрица V получается специальным построением. Причём, если $r < \min\{m, n\}$, то матрицы U, V можно получить не единственным образом.

При этом, если матрица A вещественна, то и унитарные матрицы U, V можно выбрать вещественными.

Сингулярное разложение

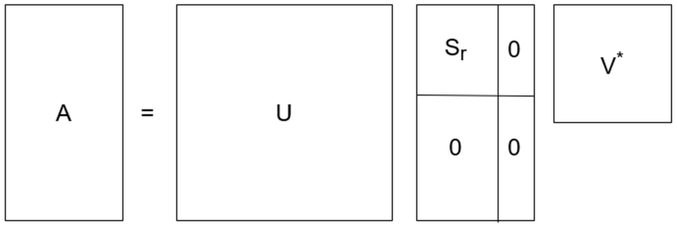
Проиллюстрируем теорему о полном SVD.

Матрица A размера $m \times n$ с рангом $\operatorname{rank} A = r$.

Матрицы U размера $m \times m$ и V размера $n \times n$ унитарны.

$$A = U \Sigma V^*, \quad \Sigma = \begin{pmatrix} S_r & \mathbb{O}_{r \times (n-r)} \\ \mathbb{O}_{(m-r) \times r} & \mathbb{O}_{(m-r) \times (n-r)} \end{pmatrix}.$$

Пусть $m > n$, то есть матрица данных такова, что объектов больше, чем признаков.



Из иллюстрации ясно, что нет необходимости хранить те столбцы матриц U и V, которые умножаются на нули из центральной матрицы. Это является мотивацией для неполного SVD (англ. economy SVD).

Сингулярное разложение

Теорема (теорема о неполном SVD). Пусть A является матрицей размера $m \times n$ с рангом $\operatorname{rank} A = r$. Тогда найдутся матрицы U_r размера $m \times r$ и V_r размера $n \times r$, удовлетворяющие равенствам

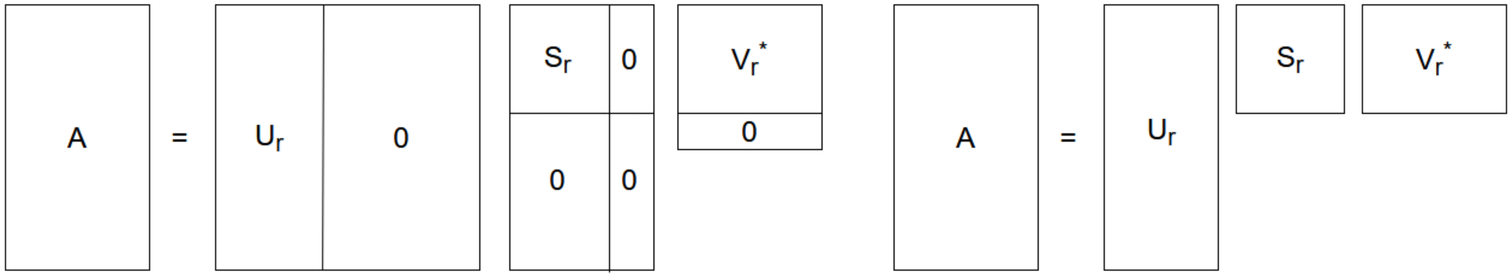
$$U_r^* U_r = I_r, \quad V_r^* V_r = I_r, \quad \text{такие что} \quad A = U_r S_r V_r^*,$$

где $S_r = \operatorname{diag}(\sigma_1, \ \ldots, \ \sigma_r)$, $\sigma_1 \geq \ldots \geq \sigma_r > 0$.

При этом если матрица A вещественна, то и матрицы U_r, V_r можно выбрать вещественными.

В теореме о неполном SVD матрицы U_r, V_r находятся единственным образом:
U_r — это первые r столбцов матрицы U,
$V_r = A^* U_r S_r^{-1}$.

Ниже проиллюстрированы случаи полного и неполного SVD, когда $m > n$.



Главные компоненты матрицы

Пусть A является матрицей размера $m \times n$ с рангом $\operatorname{rank} A = r$. Рассмотрим неполное SVD $A = U_r S_r V_r^*$.

Пусть $u_1, \ldots, u_r$ вектора-столбцы матрицы U_r, $v_1, \ldots, v_r$ вектора-столбцы матрицы V_r.

Пару векторов-столбцов (u_j, v_j), $j = 1, \ldots, r$, называют парой **левых и правых сингулярных** векторов матрицы A, ассоциированных с сингулярным значением σ_j.

Пару векторов (v_1, u_1) соответствующих наибольшему сингулярному числу σ_1, также называют (первой) **главной компонентой** матрицы A. Аналогично определяются i-ые **главные компоненты** матрицы A.

Переписав вид SVD можно представить матрицу A как специальную сумму главных компонент.

Пусть $S_r = S_{r,1} + S_{r,2} + \ldots + S_{r,r}$, где $S_{r,i} = \operatorname{diag}(0, \ldots, 0, \sigma_i, 0, \ldots, 0)$. Тогда

$$A = U_r S_r V_r^* = \sum_{i=1}^{r} U_r S_{r,i} V_r^* = u_1 \sigma_1 v_1{}^* + u_2 \sigma_2 v_2{}^* + \ldots + u_r \sigma_r v_r{}^*.$$

Слагаемыми являются матрицы ранга 1.

Экстремальное свойство SVD: обозначения

Существует множество различных разложений матрицы. Однако, SVD обладает некоторым экстремальным свойством.

Определение. Нормой Фробениуса (или также говорят нормой Гильберта-Шмидта) матрицы $A \in \mathbb{C}^{m\times n}$ называют норму

$$\|A\|_F = \left(\sum_{j=1}^{m}\sum_{k=1}^{n}|a_{j,k}|^2\right)^{1/2}.$$

Теорема (норма Фробениуса равна ℓ_2-норме сингулярных чисел). Пусть $A \in \mathbb{C}^{m\times n}$ с рангом $\operatorname{rank} A = r$. Тогда

$$\|A\|_F = \left(\sum_{j=1}^{r}\sigma_j^2\right)^{1/2},$$ где $\sigma_1, \ldots, \sigma_r$ сингулярные числа матрицы A.

Доказательство. Пусть $B = A\,A^*$. Сначала отметим, что норма Фробениуса матрицы B согласуется со следом матрицы B.

$$\operatorname{Tr} B = \sum_{j=1}^{m} b_{j,j} = \sum_{j=1}^{m}\left(\sum_{k=1}^{n} a_{j,k}\,\overline{a_{j,k}}\right) = \sum_{j=1}^{m}\sum_{k=1}^{n}|a_{j,k}|^2 = \|A\|_F^2,$$

Как известно, след матрицы $\operatorname{Tr} B$ является суммой собственных чисел матрицы B. Причём собственные числа матрицы B являются квадратами сингулярных чисел матрицы A

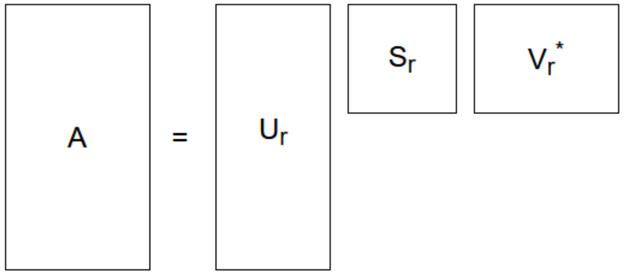
$\lambda_j = \sigma_j^2,\ \ j = 1, \ldots, m.$ Значит $\|A\|_F = \sqrt{\operatorname{Tr} B} = \sqrt{\sum_{j=1}^{m}\sigma_j^2} = \sqrt{\sum_{j=1}^{r}\sigma_j^2}$, так как $\sigma_{r+1} = \ldots = \sigma_m = 0.$•

Экстремальное свойство SVD: низкоранговые приближения

Под **наилучшим низкоранговым приближением** матрицы A понимается матрица того же размера, что и матрица A, но которая имеет меньший ранг и ближе к исходной матрице по некоторой метрике среди всех других матриц того же ранга.

Рассмотрим метрику, порождённую нормой Фробениуса. Тогда именно SVD решает задачу низкорангового приближения в этой метрике.

SVD матрицы A имеет вид : $A = U \Sigma V^* = U_r S_r V_r^*$



где U, V унитарные матрицы размера $m \times m$, $n \times n$, матрицы $U_r = (u_1, \ldots, u_r)$, $V_r = (v_1, \ldots, v_r)$ получены из U, V сохранением первых r столбцов, σ_j сингулярные числа матрицы A, расположенные по убыванию на диагонали матрицы S_r, $S_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$.

Если матрице S_r сохранить только d наибольших сингулярных чисел, $d \le r$, то так и будет получено оптимальное приближение матрицы A с помощью матрицы ранга d.

Экстремальное свойство SVD: низкоранговые приближения

Обозначим за $S_r(d)$ матрицу размера $r \times r$ вида $S_r(d) = \mathrm{diag}\,\{\sigma_1, \ldots, \sigma_d, 0, \ldots, 0\}$.

SVD матрицы A имеет вид : $A = U_r\, S_r\, V_r^*$

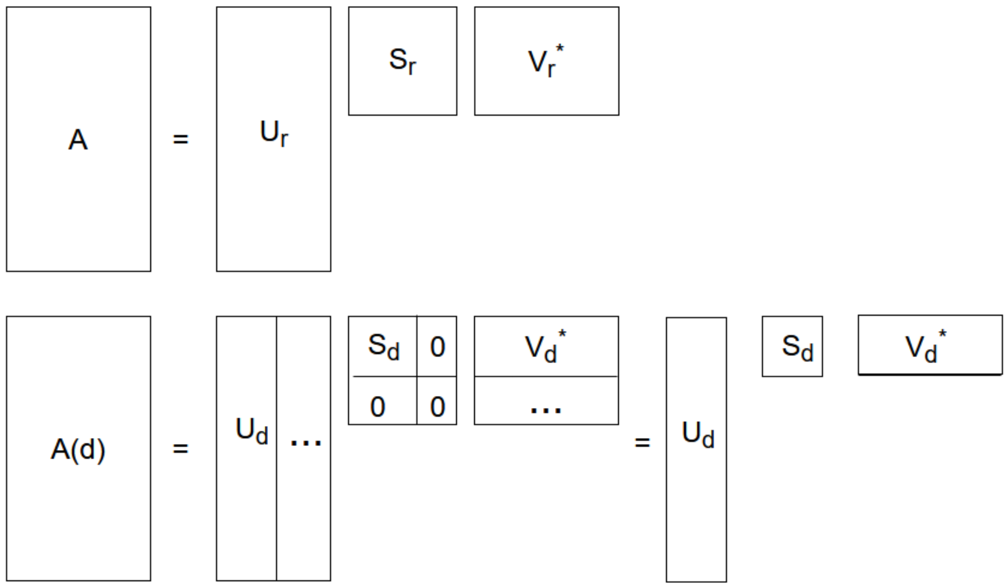
Обозначим за $A(d)$ матрицу : $A(d) = U_r\, S_r(d)\, V_r^* = U_d\, S_d\, V_d{}^*$,

где U_d, V_d — это матрицы, составленные из первых d столбцов матриц U_r, V_r,

$S_d = \mathrm{diag}\,\{\sigma_1, \ldots, \sigma_d\}$.

Матрицу $A(d)$ ещё называют сокращённым SVD (англ.truncated SVD). Она решает задачу низкорангового приближения.

Ниже проиллюстрировано различие между неполным и сокращённым SVD.



Экстремальное свойство SVD: низкоранговые приближения

Теорема (о низкоранговых приближениях). Пусть $A \in \mathbb{C}^{m\times n}$ матрица с рангом $\operatorname{rank} A = r$ и сингулярными числами $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = 0$. Тогда для любого числа d матрица сокращённого SVD вида

$$A(d) = U_r\, S_r(d)\, V_r^*$$

обеспечивает наилучшее приближение матрицы A с помощью множества всех матриц ранга не выше d в норме Фробениуса, при этом величина ошибка равна

$$\|A - A(d)\|_F^2 = \sum_{j=d+1}^{r} \sigma_j^2.$$

Иными словами,

$$\underset{\hat{A},\ \operatorname{rank}(\hat{A})=d}{\operatorname{argmin}} \left\|A - \hat{A}\right\|_F = A(d) = U_r\, S_r(d)\, V_r^*, \quad \text{где } S_r(d) \text{ — матрица размера } r\times r, \quad S_r(d) = \operatorname{diag}\{\sigma_1, \ \ldots, \sigma_d, 0, \ \ldots, 0\}.$$

Более того, известно, что SVD даёт ещё и оптимальное приближение в операторной норме (или спектральной норме).

$$\underset{\hat{A},\ \operatorname{rank}(\hat{A})=d}{\operatorname{argmin}} \left\|A - \hat{A}\right\|_2 = A(d) = U_r\, S_r(d)\, V_r^*, \quad \text{причём } \left\|A - \hat{A}\right\|_2 = \sigma_{d+1}.$$

Эта теорема также известна как теорема Экарта-Юнга.