1. Переход к новым признакам

Как раньше, $\mathbb{X} \in \mathbb{R}^{n \times p}$ — матрица данных, столбцы — вектора признаков, строки — индивиды (наблюдения).

$Z_i = \mathbb{X} A_i \in \mathbb{R}^n$, $i = 1, \ldots, d$ — новые признаки (было в прошлый раз), A_i — коэффициенты линейной комбинации. В матричном виде

$$\mathbb{Z} = \mathbb{X}\mathbb{A} \in \mathbb{R}^{n \times d},$$

где столбцы матрицы $\mathbb{A}$ — вектора из коэффициентов линейных комбинаций.

Пусть новые признаки Z_i — ортогональные и их количество равно рангу d матрицы данных $\mathbb{X}$. Тогда эти вектора составляют ортогональный базис пространства столбцов (пространства признаков).

Нормируем их $Q_i = \frac{Z_i}{\|Z_i\|}$, получим ортонормированный базис.

В ортонормированном базисе удобно вычислять координаты вектора через скалярные произведения с базисными векторами.

Отступление: *Если $U_1, \ldots, U_d$ — ортономированный базис в $\mathbb{R}^d$, то $\forall\ A \in \mathbb{R}^d$ раскладывается по ортонормированному базису:*

$$A = \sum_{i=1}^{d} \langle A, U_i \rangle U_i, \text{ где } \langle A, U_i \rangle \text{ — } i\text{-ая координата вектора } A \text{ в базисе } \{U_j\}_{j=1}^d.$$

Задание: *Рассмотрим пример: пространство $\mathbb{R}^2$, соответствующий базис $(1,1)^{\mathrm{T}}/\sqrt{2}$, $(1,-1)^{\mathrm{T}}/\sqrt{2}$. Это ортонормированный базис, да? Как вычислить координаты вектора $(5,4)^{\mathrm{T}}$ в данном базисе? Нарисуйте картинку, где отметьте координаты.*

Таким образом, $\{Q_i\}_{i=1}^d$ — ортонормированный базис в пространстве $\mathrm{span}(X_1, \ldots, X_p)$.

Исходные признаки выражаются через новые (есть пространство, в нём базис, каждый вектор этого пространства раскладывается по базису):

$$\forall\, i = 1, \ldots, p \quad X_i = \sum_{j=1}^{d} \langle X_i, Q_j \rangle Q_j. \tag{1}$$

Если ввести матрицу факторных весов (факторных нагрузок) $\mathbb{F} = \{f_{ij}\}_{i=1,j=1}^{p,d} = [F_1, : \ldots : F_d] \in \mathbb{R}^{p \times d}$, то можно записать разложение (1) в матричном виде:

$$\mathbb{X} = \sum_{j=1}^{d} Q_j F_j^{\mathrm{T}} = \mathbb{Q}\mathbb{F}^{\mathrm{T}}. \tag{2}$$

(Здесь строки матрицы $\mathbb{F}$ — коэффициенты разложения исходных признаков по новым, ортонормированным, поэтому она транспонированная.)

Важно: разложение матрицы данных тесно связано с введением новых признаков. После нормировки $P_j = \frac{F_j}{\|F_j\|}$, получим

$$\mathbb{X} = \sum_{j=1}^{d} \sigma_j Q_j P_j^{\mathrm{T}} = \mathbb{Q}\mathbf{\Sigma}\mathbb{P}^{\mathrm{T}}, \text{ где } \mathbf{\Sigma} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_d\}. \tag{3}$$

Что за матрица $Q_j P_j^{\mathrm{T}} \in \mathbb{R}^{n\times p}$? Можно заметить, что $\mathrm{rank}(Q_j P_j^{\mathrm{T}}) = 1$; действительно, столбцы матрицы пропорциональны; то же самое для строк. Таким образом, у нас была исходная матрица ранга d, а мы превратили её в сумму d элементарных матриц ранга 1 $\mathbb{X} = \sum\limits_{j=1}^{d} \mathbb{X}^{(j)}$, где $\mathbb{X}^{(j)} = \sigma_j Q_j P_j^{\mathrm{T}}$.

Вывод: для нахождения новых признаков, которые самые лучшие в каком-то смысле, нам понадобятся матричные разложения. Далее мы займемся изучением самого лучшего матричного разложения, но сначала обсудим несколько фактов из линейной алгебры.

2. Пара фактов из линейной алгебры

1. *Унитарная матрица* $\mathbb{U}$ — ортогональная матрица в комплексном случае. Ее свойства:

   - $\mathbb{U}$ — квадратная матрица: $\mathbb{U}^{\mathrm{T}} = \mathbb{U}^{-1}$.
   - Столбцы $\mathbb{U}$ ортонормированы.
   - Строки $\mathbb{U}$ ортонормированы.[1]
   - Умножение на матрицу $\mathbb{U}$ означает *поворот* или *отражение*.
   - При умножении на матрицу $\mathbb{U}$ (и на $\mathbb{U}^{\mathrm{T}}$) не меняются нормы векторов и углы между векторами. Пусть есть вектора Y, Z, после умножения на матрицу $\mathbb{U}$ получим $\widetilde{Y} = \mathbb{U}Y$, $\widetilde{Z} = \mathbb{U}Z$ и $\|Y\| = \|\widetilde{Y}\|$, $\|Z\| = \|\widetilde{Z}\|$, $< Y, Z > = < \widetilde{Y}, \widetilde{Z} >$ (почему это означает равенство углов между векторами?).

   Пример 1. $\mathbb{U} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$ — *матрица поворота на угол ϕ.*

[1] 2 пункт эквивалентен 3. *Почему?* Если матрица ортогональная, то и транспонированная к ней тоже ортогональная (следует из пункта 1).

Обычно, унитарная матрица строится из ортонормированного базиса, который составляет столбцы матрицы. (**Задание**. Проверьте, что в примере с поворотом на ϕ это так, т.е., матрица составлена из базисных векторов.)

2. Пусть $\{P_i\}_{i=1}^r$ — система независимых векторов, рассмотрим линейную оболочку $\mathcal{L}_r = \mathrm{span}\{P_1, \ldots, P_r\}$ в $\mathbb{R}^L$, $\mathbf{\Pi} : \mathbb{R}^L \to \mathcal{L}_r$ — ортогональный проектор на $\mathcal{L}_r$ (он сопоставляет вектору ближайшую точку из подпространства, что делается опусканием перпендикуляра). Матрица Π:

$$\mathbf{\Pi} = \mathbb{P}(\mathbb{P}^{\mathrm{T}}\mathbb{P})^{-1}\mathbb{P}^{\mathrm{T}}.$$

Пусть $\{P_i\}_{i=1}^r$ — ортонормированный базис $\mathcal{L}_r$.

Задание: какой вид тогда имеет матрица проектора?

$$\mathbb{P}^{\mathrm{T}}\mathbb{P} = \mathbb{I}_{r\times r} = \begin{pmatrix} 1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{r\times r}.$$

Поэтому $\mathbf{\Pi} = \mathbb{P}\mathbb{P}^{\mathrm{T}}$.

Задание. Для трехмерного вектора проверить, что его проекция на плоскость первых двух координат вычисляется по такой формуле.

Глава 1

Сингулярное разложение (SVD — Singular Value Decomposition)

Итак, приступаем к изучению самого лучшего, самого красивого, самого оптимального, самого симметричного разложения матриц.

1.1. Как строится сингулярное разложение

Мы сейчас поменяем обозначения и будем раскладывать транспонированную матрицу данных $\mathbb{Y} = \mathbb{X}^{\mathrm{T}}$, представляйте $\mathbb{Y}$ как широкую матрицу с небольшим числом строк и большим числом столбцов.

Пусть L — число признаков, K — количество индивидов, $\mathbb{Y} = \mathbb{X}^{\mathrm{T}} \in \mathbb{R}^{L\times K}$ — ненулевая матрица. Обозначим $\mathbb{S} = \mathbb{Y}\mathbb{Y}^{\mathrm{T}} \in \mathbb{R}^{L\times L}$ — симметричная неотрицательно определённая матрица. По определению,

$$\mathbb{S}U_i = \lambda_i U_i, \text{ где}$$

$\{U_i\}_{i=1}^{L}$ — ортонормированный набор из собственных векторов матрицы $\mathbb{S}$,
$\lambda_1 \geq \ldots \geq \lambda_L \geq 0$ — собственные числа матрицы $\mathbb{S}$.[1]

Пусть $d = \operatorname{rank}\mathbb{Y} = \operatorname{colrank}\mathbb{Y} = \operatorname{rowrank}\mathbb{Y}$. Знаем, что $d \leq \min(L, K)$.

Предложение 1. *1.* $d = \operatorname{rank}\mathbb{Y}\mathbb{Y}^{\mathrm{T}}$.

2. $\lambda_d > 0$; $\lambda_i = 0$ *при* $i > d$.[2]

3. $\{U_i\}_{i=1}^{d}$ *образуют ортонормированный базис* $\operatorname{colspan}\mathbb{Y}$.

Введём вектор

$$V_i \overset{def}{=} \frac{\mathbb{Y}^{\mathrm{T}}U_i}{\sqrt{\lambda_i}} \in \mathbb{R}^k,\ \ i = 1, \ldots, d.$$

Предложение 2. *1.* $\{V_i\}_{i=1}^{d}$ *— ортонормированная система векторов.*

[1] Неотрицательные, так как матрица $\mathbb{S}$ неотрицательно определена.
[2] Упорядочили собственные числа: первые d строго положительные, а остальные все нули.

2. *V_i — собственные вектора $\mathbb{Y}^{\mathrm{T}}\mathbb{Y}$, соответствующие тем же собственным числам λ_i. Остальные собственные вектора $\mathbb{Y}^{\mathrm{T}}\mathbb{Y}$ соответствуют нулевым собственным числам.*

3. $U_i = \dfrac{\mathbb{Y}V_i}{\sqrt{\lambda_i}}$.

4. $\mathbb{Y} = \sum\limits_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$ — ***SVD (Сингулярное разложение матрицы).***[3][4]

Что здесь считать новыми признаками, если $\mathbb{Y} = \mathbb{X}^{\mathrm{T}}$? V_i,[5] так как $U_i \in \mathbb{R}^L$, $V_i \in \mathbb{R}^K$.

- U_i — ортонормированный базис в пространстве столбцов.
- V_i — ортонормированный базис в пространстве строк.
- $\dfrac{\lambda_i}{\sum_{i=1}^{d} \lambda_i}$ — вклад i-ого признака.

Терминология: $\sqrt{\lambda_i}$ — *сингулярные числа* матрицы $\mathbb{Y}$, U_i — *левые сингулярные вектора*, V_i — *правые сингулярные вектора.*

Тройка $(\sqrt{\lambda_i}, U_i, V_i)$ называется i-ой собственной тройкой сингулярного разложения.

Замечание 1. *Сингулярное разложение — единственное разложение с двумя ортонормированными базисами. Оно симметрично в след. смысле: можем $\mathbb{Y}$ транспонировать, проделать всё то же самое, а поменяются местами только U_i и V_i.*

Задание. Транспонируйте матрицу $\mathbb{Y}$ и покажите, что, действительно, U_i и V_i поменяются местами (то, что называлось U, станет называться V, и наоборот).

Вернемся к SVD $\mathbb{Y} = \sum\limits_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^{\mathrm{T}} = \sum\limits_{i=1}^{d} \mathbb{Y}_i$.

Введем норму Фробениуса матрицы, квадрат которой равен сумме квадратов элементов матрицы. Так как U_i и V_i по норме равны 1, то можно показать, что $\|\mathbb{Y}_i\|_F^2 = \lambda_i$. А так как λ_i упорядочены по убыванию, то норма $\|\mathbb{Y}_1\|_F$ самая большая, у второй матрицы норма поменьше и т.д. А так как, напомню, разложение связано с введением новых признаков, то и первый новый признак самый важный, и т.д. по убыванию.

[3] Разложение в сумму элементарных матриц.

[4] Самый важный пункт утверждения.

[5] Они длинные :)

Также можно показать, что $\langle \mathbb{Y}_i, \mathbb{Y}_j \rangle_F = 0$ при $i \neq j$; поэтому получаем что-то вроде теоремы Пифагора: $\|\mathbb{Y}\|_F^2 = \sum\limits_{i=1}^{d} \|\mathbb{Y}_i\|_F^2 = \sum\limits_{i=1}^{d} \lambda_i$.

Отсюда становится очевидным, почему вклад i-й матрицы (и, соответственно, i-го признака) можно определить как $\dfrac{\lambda_i}{\sum_{i=1}^{d} \lambda_i}$.

Вопрос: Очевидно?

Задание: доказать, что столбцы матрицы $\mathbb{Y}_i$ состоят из проекций столбцов матрицы $\mathbb{Y}$ на подпространство $\operatorname{span} U_i$.

Отступление: SVD можно использовать для компактного хранения данных, если ранг матрицы маленький.

Задание: Посчитайте объем памяти для хранения всей матрицы или ее сингулярного разложения.

1.2. Единственность сингулярного разложения

Насколько единственно разложение SVD (оно одно существует для матрицы или нет)? Можно подумать, что разложение не единственное, так как

1. Собственные вектора не единственные, то есть если U_i — собственный вектор, то $(-U_i)$ — собственный вектор,

$$\mathbb{Y} = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^{\mathrm{T}} = \sum_{i=1}^{d} \sqrt{\lambda_i} (-U_i)(-V_i)^{\mathrm{T}}.$$

2. Пусть есть два одинаковых собственных числа $\lambda = \lambda_1 = \lambda_2$, U_1 и U_2 — два ортонормированных вектора, соответствующих собственному числу λ. Тогда любая линейная комбинация U_1 и U_2 будет являться также собственным вектором и будет соответствовать тому же собственному числу, то есть $\forall\ \alpha, \beta$: $\alpha U_1 + \beta U_2$ — с.в. с с.ч. λ.

   Задание. Доказать это.

   Таким образом, если у нас есть два одинаковых собственных числа, то они порождают подпространство размерности 2, и любой ортонормированный базис в этом подпространстве подходит нам в качестве собственного вектора.[6]

[6] Если у нас есть два одинаковых собственных числа, то мы можем брать любой базис, но сумма двух матриц постоянна, то есть она не меняется от выбора базиса.

Задание. Постройте сингулярное разложение матрицы, на диагонали которой стоит число 2, остальные нули.

Получаем, что единственности в буквальном смысле не получается. Поэтому сформулируем необходимое нам утверждение.

Предложение 3 (Единственность SVD)**.** *Пусть* $L \leq K$*. Пусть* $\mathbb{Y} = \sum\limits_{i=1}^{L} c_i P_i Q_i^{\mathrm{T}}$ *— некоторое разложение в сумму элементарных матриц (биортогональное разложение), такое что:*

1. $c_1 \geq \ldots \geq c_L \geq 0$*;*
2. $\{P_i\}_{i=1}^{L}$ *— ортонормированные,* $\{Q_i\}_{i=1}^{L}$ *— ортонормированные.*

Тогда $\mathbb{Y} = \sum\limits_{i=1}^{L} c_i P_i Q_i^{\mathrm{T}}$ *— SVD, то есть* ***любое биортогональное разложение с неотрицательными коэффициентами является сингулярным.***

Замечание 2. *В частности:*

- $c_d > 0$, $c_{d+1} = \ldots = c_L = 0$,
- $c_i^2 = \lambda_i$ *— собственные числа* $\mathbb{Y}\mathbb{Y}^{\mathrm{T}}$*,*
- P_i *— собственные вектора* $\mathbb{Y}\mathbb{Y}^{\mathrm{T}}$*,*
- Q_i *— собственные вектора* $\mathbb{Y}^{\mathrm{T}}\mathbb{Y}$*,*
- $Q_i = \frac{\mathbb{Y}^{\mathrm{T}} P_i}{\sqrt{\lambda_i}}$, $i = 1, \ldots, d$ *(*$d = \operatorname{rank} \mathbb{Y}\mathbb{Y}^{\mathrm{T}}$*).*

Задание. Является ли разложение матрицы

$$\mathbb{Y} = (1,1)^{\mathrm{T}}/\sqrt{2}(1,1,1)/\sqrt{3} + (-1,1)^{\mathrm{T}}/\sqrt{2}(1,-1,1)/\sqrt{3}$$

сингулярным? Заодно посчитайте, какая получается матрица $\mathbb{Y}$.

А это (матрица другая)

$$\mathbb{Y} = (1,1)^{\mathrm{T}}/\sqrt{2}(1,1,1)/\sqrt{3} + (-1,1)^{\mathrm{T}}/\sqrt{2}(2,-1,-1)/\sqrt{6}?$$

Если разложение сингулярное, выпишите сингулярные тройки, упорядочите их по λ_i.

1.3. Матричный вид сингулярного разложения

Можно записать двумя способами:

1. Введём $\mathbb{U}_d = [U_1 : \ldots : U_d]$, $\mathbb{V}_d = [V_1 : \ldots : V_d]$, $\mathbf{\Lambda}_d = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. Тогда
$$\mathbb{Y} = \mathbb{U}_d \mathbf{\Lambda}_d^{1/2} \mathbb{V}_d^{\mathrm{T}}.$$

2. Возьмём $\mathbb{U} = [U_1 : \ldots : U_d : U_{d+1} : \ldots : U_L]$ — ортонормированный базис в $\mathbb{R}^L$.[7]
$\mathbb{V}^{\mathrm{T}} = [V_1 : \ldots : V_d : V_{d+1} : \ldots : V_K]$ — ортонормированный базис в $\mathbb{R}^K$.
$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \ldots & \ldots & 0 \\ 0 & \ddots & & 0 & 0 \\ 0 & & \lambda_d & & 0 \\ 0 & & & \ddots & 0 \\ 0 & 0 & \ldots & & 0 \end{pmatrix} \in \mathbb{R}^{L \times K}.$$
Тогда[8]
$$\mathbb{Y} = \mathbb{U} \mathbf{\Lambda}^{1/2} \mathbb{V}^{\mathrm{T}}.$$

Задание. Записать в матричной форме предыдущий пример.

[7] $U_{d+1}, \ldots, U_L$ соответствуют нулевому собственному числу матрицы.
[8] $\mathbb{U}, \mathbb{V}$ — ортогональные матрицы.