

Конспектировала: Ермолова Марина.

## 1. Задачи из домашнего задания

### 1.1. Задача 1.13

Вспомним определения:

$$TSS = \sum_i (y_i - \bar{y})^2$$

$$RSS = \sum_i (y_i - \hat{y}_i)^2$$

$$ESS = \sum_i (\hat{y}_i - \bar{y})^2$$

1. Предположим, что после добавления точки к регрессии величина  $RSS$  уменьшится. Тогда получается, что

$$\sum_{i=1}^n (y_i - \hat{y}_i^*)^2 \leq \sum_{i=1}^{n+1} (y_i - \hat{y}_i^*)^2 < \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Из этого следует, что новая регрессия обеспечивает меньшую сумму квадратов, а значит исходная регрессия не была оптимальной. Противоречие.

$$RSS \leq RSS^*$$

2. Проведем рассуждения, аналогичные прошлому пункту: Допустим,  $TSS \leq TSS^*$

$$\sum_{i=1}^n (y_i - \bar{y}^*)^2 \leq \sum_{i=1}^{n+1} (y_i - \bar{y}_i^*)^2 < \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

Известно, что если проецировать любой вектор на  $\vec{1}$ , то получается  $\bar{y}\vec{1}$ . Но мы знаем, что при решении задачи минимизации

$$\min_a \sum (y_i - a)^2$$

$$a_{opt} = \bar{y}$$

А по нашему предположению  $\bar{y}^*$  дает минимум меньше, чем мы получаем, оптимизируя. Возникает противоречие.

3. Величина  $ESS$  может меняться как угодно. Попробуем решить по аналогии с предыдущими пунктами:

$$\sum_{i=1}^n (\hat{y}_i^* - \bar{y}^*)^2 \leq \sum_{i=1}^{n+1} (\hat{y}_i^* - \bar{y}_i^*)^2 < \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Видим, что провести аналогичные рассуждения мы не можем.

- Величина  $ESS$  может падать. Например, можно разместить точку настолько низко, чтобы прямая стала горизонтальной. Тогда все  $\hat{y}_i = \bar{y}^*$  и  $ESS = 0$ .
- Величина  $ESS$  может возрастать. Расположим  $\hat{y}_i$  на прямую, тогда величина  $\hat{y}$  не меняется, величина  $\bar{y}$  выросла, появилось новое слагаемое, из-за которого величина  $ESS$  возрастает.

## 2. Сингулярное разложение матрицы

Английское сокращение: SVD — Singular Value Decomposition. Сингулярное разложение матрицы используется при снижении размерности и не только.

### 2.1. Смысл условия $U^T U = I$

Вспомним смысл условия  $U^T U = I$ , где  $I$  — единичная матрица.

1. Порядок умножения не важен:

$$U^T U = I \Leftrightarrow U^{-1} = U^T \Leftrightarrow U U^T = I$$

2. Рассмотрим вектор-столбцы матрицы  $U$ ,  $u_1, \dots, u_n$ :

$$\begin{bmatrix} \dots & u_1^T & \dots \\ \dots & u_2^T & \dots \\ \dots & \dots & \dots \\ \dots & u_n^T & \dots \end{bmatrix} \begin{bmatrix} \vdots & \vdots & \vdots \\ u_1 & u_2 & u_n \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$u_2^T \cdot u_2 = 1$$

$$u_2^T \cdot u_7 = 0$$

Следовательно, столбцы  $u$  имеют единичную длину и ортогональны друг другу.

### 2.2. Свойства ортогональной матрицы

$$U \cdot x = \tilde{x}$$

1. Матрица  $U$  сохраняет длину вектора.

$$\tilde{x}^T \tilde{x} = (Ux)^T (Ux) = x^T U^T U x = x^T I x = x^T x$$

2. Матрица  $U$  сохраняет угол.

$$U \cdot x = \tilde{x}$$

$$U \cdot y = \tilde{y}$$

$$\cos(\tilde{x}, \tilde{y}) = \frac{\tilde{x}^T \tilde{y}}{\|\tilde{x}\| \|\tilde{y}\|} = \frac{x^T U^T U y}{\|x\| \|y\|} = \frac{x^T y}{\|x\| \|y\|} = \cos(x, y)$$

Матрица  $U$  может быть, например, поворотом или отражением.

### 2.3. Идея SVD геометрически

Пусть  $X$  — матрица переменных размера  $n \times k$ , где  $k$  — количество переменных,  $n$  — количество наблюдений. Матрица  $X$  превращает  $k$ -мерные вектора в  $n$ -мерные.

$$X \cdot a = b$$

Любое действие  $X$  при выборе удачного базиса в  $\mathbb{R}^k$  и при выборе удачного базиса в  $\mathbb{R}^n$  очень просто действует.

Допустим, в пространстве  $\mathbb{R}^k$  мы выбрали базис  $v_1, v_2, \dots, v_k$ . А в  $\mathbb{R}^n$  мы выбрали базис  $u_1, u_2, \dots, u_n$ . Оказывается, можно так подобрать базисы, что:

$$\begin{cases} Xv_1 = \lambda_1 u_1 \\ Xv_2 = \lambda_2 u_2 \\ \dots \end{cases}$$

Так как в общем случае  $n \neq k$ , то, либо в некоторые  $u$  не переходят  $v$ , либо если  $k > n$ ,  $Xv_j = 0, j > n$ .

**Пример.** Проецирование  $X : \mathbb{R}^3 \rightarrow \mathbb{R}^2$

С помощью сингулярного разложения независимо от действия можно выбрать хороший базис в  $\mathbb{R}^2$  и  $\mathbb{R}^3$ , что действие будет очень просто. Выберем базис:

$$\begin{cases} Xv_1 = u_1 \\ Xv_2 = u_2 \\ Xv_3 = 0 \end{cases}$$

**Пример:** Поворот на плоскости по часовой стрелке на 45 градусов  $X : \mathbb{R}^2 \rightarrow \mathbb{R}^2$

Выберем базис:

$$\begin{cases} Xv_1 = u_1 \\ Xv_2 = u_2 \end{cases}$$

### 2.4. Представление матрицы в виде произведения

Любую матрицу  $X$  размера  $n \times k$  можно представить в виде  $X = U \cdot \Sigma \cdot V^T$ , где:  $U^T \cdot U = I_{n \times n}$ ,  $V^T \cdot V = I_{k \times k}$ , т.е. действие  $V$  не меняет углы и расстояния, а просто меняет базис.

Матрица  $\Sigma$  диагональная, но необязательно квадратная. Например,

$$\Sigma = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Пусть  $x_V$  — столбец координат в базисе  $V$ , а  $x$  — столбец координат в исходном базисе, тогда

$$Vx_V = x$$

$$x_V = V^T x$$

## 2.5. Задача 5.14

Дана матрица  $X = U \begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} V^T$ .

Найти сингулярное разложение матриц:

1.  $X^T$
2.  $10X$
3.  $X^T X$
4.  $XX^T$

**Решение:**

$$1. X^T = \left( U \begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} V^T \right)^T = V \begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}^T U^T = V \begin{pmatrix} 7 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} U^T$$

$$2. \text{ Матрицы } U, V \text{ сохраняют длину, поэтому } \Sigma = \begin{pmatrix} 70 & 0 & 0 \\ 0 & 30 & 0 \end{pmatrix}$$

В итоге:

$$10X = U \begin{pmatrix} 70 & 0 & 0 \\ 0 & 30 & 0 \end{pmatrix} V^T$$

3.

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

$$X^T X = V \begin{pmatrix} 7 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 70 & 0 & 0 \\ 0 & 30 & 0 \end{pmatrix} V^T = V \begin{pmatrix} 49 & 0 \\ 0 & 9 \\ 0 & 0 \end{pmatrix} V^T$$

4.

$$XX^T = U \Sigma V^T V \Sigma^T U^T = U \Sigma \Sigma^T U^T$$

$$XX^T = U \begin{pmatrix} 70 & 0 & 0 \\ 0 & 30 & 0 \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} U^T = U \begin{pmatrix} 49 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{pmatrix} U^T$$

## 2.6. Алгоритм SVD разложения компьютером

1. Пусть  $X = UBV^T$ , где  $U, V$  — отражение, а  $B$  — bidiagonalная матрица.

$$B = \begin{pmatrix} a_{1,1} & a_{1,2} & 0 & \dots & 0 \\ 0 & a_{1,1} & a_{1,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & a_{n-1,n-1} & 0 \\ 0 & \dots & 0 & a_{n,n-1} & a_{n,n} \end{pmatrix}$$

На первом этапе действия алгоритма подбираются такие отражения, чтобы получить bidiagonalную матрицу. Этот этап условно точный, мы можем его выписать численно.

2. Этот этап численный. С помощью вращений  $X = U\Sigma V^T$  достигаем конечного результата с какой-то точностью.

Важно, что на компьютере при сингулярном разложении матрицы  $X$  никогда не считается  $X^T X$ .

## 2.7. SVD-разложение руками

**Задача.** Найти SVD-разложение матрицы

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}$$

**Решение.**

$$X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} = U\Sigma V^T$$

$$X^T X = V\Sigma^T \Sigma V^T = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$X X^T = U\Sigma \Sigma^T U^T = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Найдем собственные числа матрицы  $X^T X$ :

$$\det \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = 0 \rightarrow \lambda = 3, 1$$

Собственные вектора матрицы  $X^T X$  равны  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Нормированная матрица:  $V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$

Тогда

$$X^T X = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}^T$$

Собственные числа для матрицы  $X X^T$ :

$$\det \begin{pmatrix} 2-\lambda & 1 & -1 \\ 1 & 1-\lambda & 0 \\ -1 & 0 & 1-\lambda \end{pmatrix} = 0 \rightarrow \lambda = 0, 1, 3$$

Собственные вектора матрицы  $X X^T$ :  $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$

Нормированная матрица  $U$ :

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 1 & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{1} & \frac{\sqrt{6}}{1} \\ \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{1} & -\frac{\sqrt{6}}{1} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

В итоге получаем разложение исходной матрицы  $X$ :

$$X = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 1 & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{1} & \frac{\sqrt{6}}{1} \\ \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{1} & -\frac{\sqrt{6}}{1} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

### 3. Метод главных компонент (Principal Component Analysis)

#### 3.1. Задача

Сформулируем задачу. Она состоит в том, чтобы снизить размерность матрицы таким образом, чтобы несильно двигать точки. Идея: пусть у нас есть две координаты. Тогда мы возьмем двумерное облако точек. Мы хотим несильно двигая точки сделать так, чтобы точки лежали на одной прямой.

Пусть  $x_i$  — транспонированная  $i$ -ая строка в матрице  $X$  или столбец в матрице  $X^T$ . Тогда наша задача сводится к задаче оптимизации:

$$\min_{\mu, V, \lambda_i} \sum ||x_i - \mu - V\lambda_i||^2$$

Вектор  $x_i$  имеет размер  $k \times 1$ , вектор  $\mu$  имеет размер  $k \times 1$ , вектор  $\lambda_i$  имеет размер  $p \times 1$ , матрица  $V$  имеет размер  $k \times p$ .

В столбцах матрицы  $V$  находятся  $k$ -мерные ортогональные вектора, единичной длины. Вектор  $\lambda_i$  показывает веса, с которыми надо взять эти вектора, пытаясь заменить каждый  $x_i$ . Вектор  $\mu$  является серединой облака. Число  $k$  — исходная размерность, число  $p$  — размерность, до которой хотим снизить. Для визуализации обычно берут  $p = 2$ .

Предположим, что матрица  $V$  уже найдена. Оказывается, неважно, до какой размерности мы снижали,  $\mu = \frac{\sum x_i}{n}$ , а  $\lambda_i$  находится легко.

Допустим вектор  $\mu$  и матрица  $V$  найдены. Оптимальный вектор  $\lambda_1$  находим, решая задачу оптимизации:

$$\min_{\lambda_1} ||x_1 - \mu - V\lambda_1||^2$$

Это есть аналог МНК

$$\beta : \min_{\beta} ||y - X\beta||^2$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

В нашем случае:

$$\lambda_i = (V^T V)^{-1} V^T (x_i - \mu) = V^T (x_i - \mu)$$

На самом деле метод главных компонент — это выбор плоскости, а дальше МНК.

Почему  $V^T V = I$ ?

Пусть  $x_1, x_2, x_3, \dots$  — вектора наблюдений:  $x_i$  —  $i$ -ое наблюдение, в отличие от МНК, где  $x_1, x_2, x_3, \dots$  — регрессоры. Ищем подпространство, чтобы при проецировании на него получались  $\hat{x}_1, \hat{x}_2, \dots$ . Это подпространство будем задавать базисом  $v_1, \dots, v_p$ . Для удобства задаем подпространство ортонормальным базисом.

$$\begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Вектор  $\lambda_i$  мы уже нашли с помощью МНК, теперь осталось проминимизировать по  $\mu, V$ :

$$\min_{\mu, V} \|(x_i - \mu) - VV^T(x_i - \mu)\|^2 = \min_{\mu, V} Q_\mu$$

Матрица  $VV^T$  — проектор, матрица-шляпница.

Выпишем условие первого порядка:

$$\begin{aligned} dQ_\mu &= \sum_i \|(I - VV^T)(x_i - \mu)\|^2 = \sum_i d((x_i - \mu)^T (I - VV^T)^T (I - VV^T)(x_i - \mu)) = \sum_i d(x_i - \mu)^T (I - VV^T)(x_i - \mu) = \\ &= \sum_i -2(x_i - \mu)^T (I - VV^T) d\mu = 0 \end{aligned}$$

В итоге решение  $\hat{\mu} = \frac{\sum x_i}{n}$  зануляет дифференциал  $dQ$ . Но оно неединственно, можно придумать и другие.