

Конспектировала: Лиза Вахрамева.

1. О следе матрицы

Для квадратной матрицы $A \in \mathbb{R}^{n \times n}$ вводится понятие следа матрицы:

$$\text{trace}(A) = \sum_{i=1}^n A_{ii},$$

то есть след матрицы — это сумма её диагональных элементов. След матрицы обладает следующими основными свойствами, мотивирующими его введение и дальнейшее использование:

1. $\text{trace}(AB) = \text{trace}(BA)$,
2. след матрицы равен сумме корней характеристического уравнения,
3. след матрицы является скаляром: $\text{trace}(\cdot) \in \mathbb{R}$.

Покажем, что если у матрицы A есть n действительных собственных чисел $\lambda_1 \dots \lambda_n$, то $\text{trace}(A) = \sum_{i=1}^n \lambda_i$:

$$A = PDP^{-1},$$

где P — матрица, составленная из собственных векторов $v_1 \dots v_n$, соответствующих собственным числам $\lambda_1 \dots \lambda_n$, а матрица D — диагональная из собственных чисел:

$$\text{trace}(A) = \text{trace}(PDP^{-1}) = \text{trace}(DP^{-1}P) = \text{trace}(D) = \sum_{i=1}^n \lambda_i.$$

Упражнение: дана регрессия $\hat{y} = X\hat{\beta}$, построенная методом МНК по n наблюдениям, k регрессорам. Вектор наблюдений y можно спроецировать на пространство наблюдений \mathbb{R}^n , в котором строится предсказание, с помощью матрицы-шляпницы H :

$$\hat{y} = Hy.$$

Нужно найти $\text{trace}(H)$. Рассмотрим два возможных решения:

1. Содержательное.

Известно, что след H равен сумме собственных значений H . Также известно, что H проецирует вектора на линейную оболочку векторов-столбцов матрицы X , обозначим их $c_1 \dots c_k \in \mathbb{R}^n$. Так как всеобъемлющее пространство имеет размерность n , то набор $c_1 \dots c_k$ векторов можно дополнить ортогональным набором $p_1 \dots p_{n-k}$ до базиса. Все вектора $c_1 \dots c_k$ переходят сами в себя при проекции, поэтому являются собственными с собственными числами 1. Все вектора $p_1 \dots p_{n-k}$ являются ортогональными оболочке, на которую строится проекция, и при проекции переходят в ноль, поэтому они являются собственными векторами с собственными числами 0. Получилось $\lambda_1 \dots \lambda_k = 1$ и $\lambda_{k+1} \dots \lambda_n = 0$.

$$\text{trace}(H) = \sum_i \lambda_i = k$$

2. По определению.

$$\text{trace}(H) = \text{trace}(X(X^T X)^{-1} X^T) = \text{trace}((X^T X)^{-1} X^T X) = \text{trace}(I_{k \times k}) = k.$$

2. Ковариация и ковариационная матрица

Пусть $y_1, y_2 \in \mathbb{R}$ — скалярные случайные величины, $a, b \in \mathbb{R}^n$ — вектора-столбцы чисел.

Дисперсия СВ:

$$\text{Var}(y_1) = E(y_1^2) - (E(y_1))^2.$$

Выборочная дисперсия столбца чисел:

$$\text{sVar}(a) = \frac{\sum_i^n (a_i - \bar{a})^2}{n - 1}$$

квадрат длины центрированного вектора, деленный на размерность подпространства, в котором лежит вектор.

Ковариация двух СВ:

$$\text{Cov}(y_1, y_2) = E y_1 y_2 - E y_1 E y_2 = E \left((y_1 - E y_1)(y_2 - E y_2) \right).$$

Выборочная ковариация двух столбцов чисел:

$$\text{sCov}(a, b) = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{n - 1}$$

скалярное произведение центрированных векторов, деленное на размерность пространства, в котором они лежат.

Корреляция двух СВ:

$$\text{Corr}(y_1, y_2) = \frac{\text{Cov}(y_1, y_2)}{\sqrt{\text{Var}(y_1)} \sqrt{\text{Var}(y_2)}}.$$

Выборочная корреляция двух столбцов чисел:

$$\text{sCorr}(a, b) = \frac{\text{sCov}(a, b)}{\sqrt{\text{sVar}(a)} \sqrt{\text{sVar}(b)}} = \cos(a - \bar{a}, b - \bar{b}).$$

Упражнение: Записать выборочную ковариационную матрицу для X в предположении, что переменные уже центрированы. Запишем матрицу X по строкам и по столбцам:

$$X = \begin{bmatrix} \vdots & & \vdots \\ c_1 & \dots & c_k \\ \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} \dots & r_1^T & \dots \\ \vdots & & \vdots \\ \dots & r_n^T & \dots \end{bmatrix}.$$

Поймем, как выглядит $\text{sVar}(X)$:

$$\text{sVar}(X) = \begin{bmatrix} \text{sVar}(c_1) & \text{sCov}(c_1, c_2) & \dots & \text{sCov}(c_1, c_k) \\ \text{sCov}(c_2, c_1) & \text{sVar}(c_2) & \dots & \text{sCov}(c_2, c_k) \\ \vdots & & \ddots & \vdots \\ \text{sCov}(c_k, c_1) & \text{sCov}(c_k, c_2) & \dots & \text{sVar}(c_k, c_k) \end{bmatrix}$$

С учетом того что $c_1 \dots c_k$ центрированы, ковариационную матрицу можно найти следующими способами:

$$\text{sVar}(X) = \frac{X^T X}{n - 1} = \frac{\sum_i^n r_i r_i^T}{n - 1}.$$

Чтобы удостовериться в правильности ответов, рассмотрим одну из ячеек ковариационной матрицы:

$$\text{sVar}(X)_{21} = \frac{c_2^T c_1}{n - 1} = \frac{\sum_i^n r_{i2} r_{i1}}{n - 1}.$$

3. Метод главных компонент

В методе главных компонент рассматривается способ понижения размерности при минимальных потерях в разбросе данных.

$$\min_{\mu, V: V^T V = I, \lambda_i} \sum_i^n \|r_i - \mu - V \lambda_i\|,$$

$r_i \in \mathbb{R}^k$ — i -ая строка матрицы X , записанная в столбец, $\mu \in \mathbb{R}^k$ — вектор средних, $V \in \mathbb{R}^{k \times p}$ — матрица, столбцы которой ортонормальны: $V^T V = I$, $\lambda_i \in \mathbb{R}^p$, k — исходная размерность, p — желаемая.

Предположим, что μ и V найдены, выразим λ_i . Заметим, что для фиксированного i задача превращается в аналогичную задаче поиска коэффициентов линейной регрессии, поэтому можно сразу выписать ответ:

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

Далее будем считать, что переменные заранее центрированы, поэтому положим $\mu = 0$. Подставим выражение для λ_i в исходную задачу:

$$Q = \|r_i - V V^T r_i\| \rightarrow \min_{V: V^T V = I}.$$

Распишем:

$$Q = \sum_i^n (r_i - V V^T r_i)^T (r_i - V V^T r_i) = \sum_i^n r_i^T (I - V V^T) (I - V V^T) r_i.$$

Заметим, что $V V^T$ — это матрица-шляпница, проецирующая r_i на линейную оболочку из $v_1 \dots v_p$ столбцов матрицы V :

$$H = V(V^T V)^{-1} V^T = V V^T.$$

Тогда матрица $I - V V^T$ является матрицей, проецирующей вектора r_i на ортогональное дополнение к пространству, натянутому на оболочку $v_1 \dots v_p$. Проекция на подпространство L вектора $x \in L$ уже лежащего в этом подпространстве, равна x , поэтому $(I - V V^T)(I - V V^T)x = (I - V V^T)x \quad \forall x$. Получили:

$$Q = \sum_i^n r_i^T (I - V V^T) r_i.$$

Раскроем скобки:

$$Q = \sum_i^n r_i^T r_i - r_i^T V V^T r_i \rightarrow \min_{V: V^T V = I}.$$

Перейдем к эквивалентной задаче:

$$Q = r_i^T V V^T r_i \rightarrow \max_{V: V^T V = I}.$$

Так как $Q \in \mathbb{R}$ — скаляр, то можно записать:

$$\begin{aligned} Q &= \text{trace}\left(\sum_i^n r_i^T V V^T r_i\right) = \text{trace}(V V^T \sum_i^n r_i r_i^T) \\ &= \text{trace}(V V^T (n-1) \text{sVar}(X)) = (n-1) \text{trace}(V \text{sVar}(X) V^T) = (n-1) \sum_j^p v_j^T \text{sVar}(X) v_j \end{aligned}$$

Перейдем к эквивалентной задаче:

$$Q' = \sum_j^p v_j^T S v_j \rightarrow \max,$$

где $S = \text{sVar}(X)$. Q' интерпретируется как выборочная дисперсия линейной комбинации столбцов матрицы X , взятых с весами вектора v_j .

Упражнение. Рассмотрим матрицу $X \in \mathbb{R}^{n \times 2}$ и ее ковариационную матрицу:

$$X = \begin{bmatrix} \vdots & \vdots \\ c_1 & c_k \\ \vdots & \vdots \end{bmatrix},$$

$$\text{sVar}(X) = \begin{bmatrix} 5 & -1 \\ -1 & 16 \end{bmatrix}.$$

Будем искать выборочную дисперсию $z = 3c_1 + 6c_2$.

$$\text{Var}(3c_1 + 6c_2) = 9 \text{Var}(c_1) + 36 \text{Var}(c_2) + 18 \text{Cov}(c_1, c_2) = 9 \cdot 5 + 36 \cdot 16 + 18 \cdot (-1) = 603.$$

$$\begin{bmatrix} 3 & -6 \end{bmatrix} \begin{bmatrix} 5 & -1 \\ -1 & 16 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 9 \cdot 5 + 36 \cdot 16 + 18 \cdot (-1) = 603.$$

Получилось, что выборочную дисперсию можно искать по формуле для дисперсии суммы, а можно через построение квадратичной формы.

Можно снова переписать оптимизируемый функционал:

$$V = \begin{bmatrix} \vdots & \vdots \\ v_1 & \dots & v_p \\ \vdots & \vdots \end{bmatrix},$$

$$\sum_j^p \text{sVar}(X v_j) \rightarrow \max_{V: V^T V = I}.$$

Упражнение: Найти главную компоненту X .

$$X = \begin{bmatrix} 3 & 5 \\ 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Сначала нужно центрировать и нормировать данные:

$$\begin{bmatrix} 3 & 5 \\ 2 & 1 \\ 1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 \\ 0 & -2 \\ -1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 \\ 0 & -1 \\ -1 & 0 \end{bmatrix} = X'$$

С помощью метода главных компонент будем искать вектор $v_1 \in \mathbb{R}^2$ — веса, с которыми нужно взять линейную комбинацию столбцов матрицы X , чтобы получить первую главную компоненту. Введем обозначение:

$$v_1 = \begin{bmatrix} \alpha \\ \beta \end{bmatrix},$$

тогда требования на ортонормированность базиса из векторов v в случае с одной главной компонентой — только нормированность) можно записать так:

$$\|v_1\| = 1 \rightarrow \alpha^2 + \beta^2 = 1.$$

Запишем оптимизируемый функционал:

$$Q = \text{sVar} \left(\alpha \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right) \rightarrow \max_{\alpha, \beta: \alpha^2 + \beta^2 = 1}$$

Чтобы посчитать значение этого функционала, нужно посчитать ковариационную матрицу X' :

$$\text{sVar}(X') = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}.$$

Тогда

$$Q = \alpha^2 + \beta^2 + 2\alpha\beta\frac{1}{2} = 1 + \alpha\beta.$$

Решением задачи:

$$\begin{cases} \alpha\beta \rightarrow \max \\ \alpha^2 + \beta^2 = 1 \end{cases}$$

являются точки $\alpha = \beta = \frac{1}{\sqrt{2}}$ и $\alpha = \beta = -\frac{1}{\sqrt{2}}$. Знак здесь задает только ориентацию вектора, поэтому можно рассматривать любую из точек.

Теперь можно записать главную компоненту:

$$pc_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

Поиск нескольких главных компонент можно осуществлять, последовательно решая задачи:

$$\begin{cases} \text{sVar}(Xv_1) \rightarrow \max \\ \|v_1\| = 1 \end{cases}$$

$$\begin{cases} \text{sVar}(Xv_2) \rightarrow \max \\ \|v_2\| = 1 \\ v_2^T v_1 = 0 \end{cases}$$

$$\begin{cases} \text{sVar}(Xv_3) \rightarrow \max \\ \|v_3\| = 1 \\ v_3^T v_1 = 0 \\ v_3^T v_2 = 0 \end{cases}$$

Упражнение: Если $V^T V = I$, то V сохраняет длины.

$$\|Va\| = a^T V^T V a = \|a\|.$$

Упражнение: Если $\|a\| = 1$, то

$$\max_a a^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} a = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

4. Связь с SVD-разложением

Рассмотрим поиск первой главной компоненты:

$$\max_{v_1: \|v_1\|=1} \text{sVar}(Xv_1) = \max_{v_1: \|v_1\|=1} v_1^T \frac{X^T X}{n-1} v_1 = \max_{v_1: \|v_1\|=1} v_1^T X^T X v_1$$

SVD-разложение:

$$\begin{aligned} X &= U \Sigma V^T \\ X^T X &= V \Sigma^T \Sigma V^T \end{aligned}$$

Подставим в оптимизируемый функционал:

$$\max_{v_1: \|v_1\|=1} v_1^T V \Sigma^T \Sigma V v_1.$$

Так как $V^T V = I$, то V сохраняет длины, поэтому $\|V v_1\| = \|v_1^T V\| = 1 \iff \|v_1\| = 1$, а значит можно переписать задачу:

$$\max_{v'_1: \|v'_1\|=1} v'_1{}^T \Sigma^T \Sigma V v'_1 = \max_{v'_1: \|v'_1\|=1} v'_1{}^T \begin{bmatrix} \lambda_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda_n \end{bmatrix} v'_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

здесь $\lambda_1 \dots \lambda_n$ — собственные числа матрицы $\Sigma^T \Sigma$, отсортированные по убыванию.

Сейчас мы нашли v'_1 , но нужно найти v_1 .

$$V^T v_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$V V^T v_1 = v_1 = V \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Получилось, что искомый столбец v_1 для PCA — это столбец матрицы V из SVD разложения X , соответствующий первому по величине значению в матрице Σ ($\Sigma_{ii} = \sqrt{\lambda_i}$).

$$pc_1 = X v_1 = (U \Sigma V^T) v_1 = U \Sigma \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sqrt{\lambda_1} u_1.$$

Это верно для всех рассматриваемых главных компонент:

$$pc_j = \sqrt{\lambda_j} u_j.$$