

Линейные системы

В задаче наименьших квадратов (aka линейной регрессии) мы имеем измерения $X \in \mathbb{R}^{m \times n}$ и $y \in \mathbb{R}^m$ и ищем вектор $\theta \in \mathbb{R}^n$ такой, что $X\theta$ близок к y. Близость определяется как сумма квадратов разностей:

$$\sum_{i=1}^m (x_i^\top \theta - y_i)^2 \qquad \|X\theta - y\|_2^2 \to \min_{\theta \in \mathbb{R}^n} \qquad X\theta^* = y$$



Рис. 1: Illustration of linear system aka least squares

Moore--Penrose inverse

Если матрица X относительно мала, мы можем записать и вычислить точное решение:

$$\theta^* = (X^\top X)^{-1} X^\top y = X^\dagger y,$$



Moore--Penrose inverse

Если матрица X относительно мала, мы можем записать и вычислить точное решение:

$$\theta^* = (X^\top X)^{-1} X^\top y = X^\dagger y,$$

где X^\dagger называется псевдо-обратной матрицей. Однако, этот подход возводит в квадрат число обусловленности задачи, что может быть проблемой для больших и плохо обусловленных задач.



Moore--Penrose inverse

Если матрица X относительно мала, мы можем записать и вычислить точное решение:

$$\theta^* = (X^\top X)^{-1} X^\top y = X^\dagger y,$$

где X^\dagger называется псевдо-обратной матрицей. Однако, этот подход возводит в квадрат число обусловленности задачи, что может быть проблемой для больших и плохо обусловленных задач.

QR разложение

Для любой матрицы $X \in \mathbb{R}^{m \times n}$ существует QR разложение:

$$X = Q \cdot R$$
,

♥ O 0 4

Moore--Penrose inverse

Если матрица X относительно мала, мы можем записать и вычислить точное решение:

$$\theta^* = (X^\top X)^{-1} X^\top y = X^\dagger y,$$

где X^\dagger называется псевдо-обратной матрицей. Однако, этот подход возводит в квадрат число обусловленности задачи, что может быть проблемой для больших и плохо обусловленных задач.

QR разложение

Для любой матрицы $X \in \mathbb{R}^{m \times n}$ существует QR разложение:

$$X = Q \cdot R$$

где Q - ортогональная матрица (ее столбцы ортогональные единичные векторы) и R - верхняя треугольная матрица. Важно отметить, что поскольку $Q^{-1} = Q^{\top}$, мы имеем:

$$QR\theta = y \longrightarrow R\theta = Q^{\top}y$$

Теперь процесс нахождения θ состоит из двух шагов:

1. Найдите QR разложение X.

₩ ೧ €

Moore--Penrose inverse

Если матрица X относительно мала, мы можем записать и вычислить точное решение:

$$\theta^* = (X^\top X)^{-1} X^\top y = X^\dagger y,$$

где X^\dagger называется псевдо-обратной матрицей. Однако, этот подход возводит в квадрат число обусловленности задачи, что может быть проблемой для больших и плохо обусловленных задач.

QR разложение

Для любой матрицы $X \in \mathbb{R}^{m \times n}$ существует QR разложение:

$$X = Q \cdot R$$

где Q - ортогональная матрица (ее столбцы ортогональные единичные векторы) и R - верхняя треугольная матрица. Важно отметить, что поскольку $Q^{-1}=Q^{\top}$, мы имеем:

$$QR\theta = y \longrightarrow R\theta = Q^{\top}y$$

Теперь процесс нахождения θ состоит из двух шагов:

- 1. Найдите QR разложение X.
- 2. Решите треугольную систему $R\theta = Q^{\top}y$, которая треугольная и, следовательно, легко решаемая.

Разложение Холецкого

Для любой положительно определенной матрицы $A \in \mathbb{R}^{n \times n}$ существует разложение Холецкого:

$$X^{\top}X = A = L^{\top} \cdot L,$$

где L - нижняя треугольная матрица. Мы имеем:

$$L^{\top}L\theta = y \longrightarrow L^{\top}z_{\theta} = y$$

Теперь процесс нахождения θ состоит из двух шагов:

- 1. Найдите разложение Холецкого $X^{\top}X$.
- Обратите внимание, что в этом случае ошибка пропорциональна квадрату числа обусловленности.

♥ ೧ 0

Разложение Холецкого

Для любой положительно определенной матрицы $A \in \mathbb{R}^{n \times n}$ существует разложение Холецкого:

$$X^{\top}X = A = L^{\top} \cdot L,$$

где L - нижняя треугольная матрица. Мы имеем:

$$L^{\top}L\theta = y \longrightarrow L^{\top}z_{\theta} = y$$

Теперь процесс нахождения θ состоит из двух шагов:

- 1. Найдите разложение Холецкого $X^{\top}X$.
- 2. Найдите $z_{\theta} = L \theta$ путем решения треугольной системы $L^{\top} z_{\theta} = y$

Обратите внимание, что в этом случае ошибка пропорциональна квадрату числа обусловленности.

♥ ೧ 0

Разложение Холецкого

Для любой положительно определенной матрицы $A \in \mathbb{R}^{n \times n}$ существует разложение Холецкого:

$$X^{\top}X = A = L^{\top} \cdot L,$$

где L - нижняя треугольная матрица. Мы имеем:

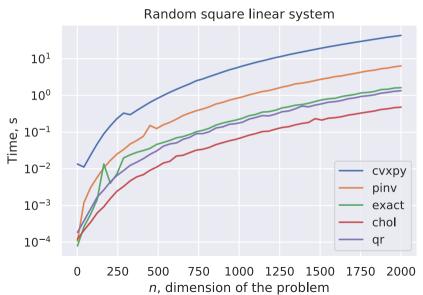
$$L^{\top}L\theta = y \longrightarrow L^{\top}z_{\theta} = y$$

Теперь процесс нахождения θ состоит из двух шагов:

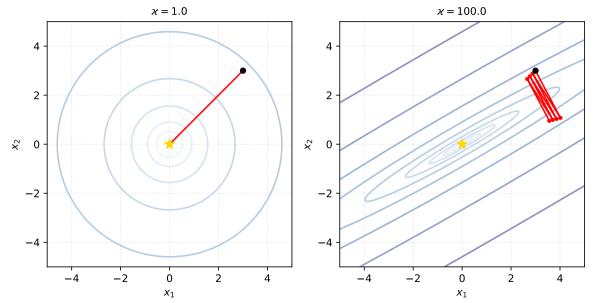
- 1. Найдите разложение Холецкого $X^{\top}X$.
- 2. Найдите $z_{\theta} = L \theta$ путем решения треугольной системы $L^{\top} z_{\theta} = y$
- 3. Найдите θ путем решения треугольной системы $L\theta=z_{\theta}$

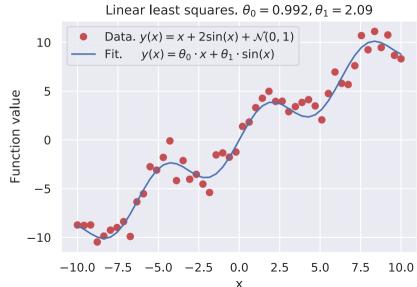
Обратите внимание, что в этом случае ошибка пропорциональна квадрату числа обусловленности.

⊕ 0 @



Число обусловленности и





Собственные векторы и собственные значения





Что такое собственный вектор?

ullet Определение. Вектор x
eq 0 называется собственным вектором квадратной матрицы A, если существует число λ такое, что

$$Ax = \lambda x$$
.



Что такое собственный вектор?

• Определение. Вектор $x \neq 0$ называется **собственным вектором** квадратной матрицы A, если существует число λ такое, что

$$Ax = \lambda x$$
.

• Число λ называется **собственным значением**. Иногда используется термин **собственная пара**.



Что такое собственный вектор?

ullet Определение. Вектор x
eq 0 называется **собственным вектором** квадратной матрицы A, если существует число λ такое, что

$$Ax = \lambda x$$
.

- ullet Число λ называется **собственным значением**. Иногда используется термин **собственная пара**.
- ullet Поскольку $A-\lambda I$ должна иметь нетривиальное ядро, собственные значения являются корнями характеристического полинома

$$\det(A - \lambda I) = 0.$$



Спектральное разложение

Если матрица A размера $n \times n$ имеет n собственных векторов $s_i, i=1,\dots,n$:

$$As_i = \lambda_i s_i,$$

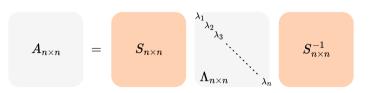
то это можно записать как

$$AS = S\Lambda, \quad \text{где} \quad S = (s_1, \dots, s_n), \quad \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n),$$

или эквивалентно

$$A = S\Lambda S^{-1}$$
.

Это называется **спектральным разложением** матрицы. Матрицы, которые могут быть представлены в виде спектрального разложения, называются **диагонализируемыми**.





Вычисление собственных значений и собственных векторов

Как вычислить собственные значения и собственные векторы?

Существуют два типа задач на собственные значения:

• вычисление всех собственных значений и собственных векторов или полного спектра (требуются все собственные значения и собственные векторы)



Вычисление собственных значений и собственных векторов

Как вычислить собственные значения и собственные векторы?

Существуют два типа задач на собственные значения:

- вычисление всех собственных значений и собственных векторов или полного спектра (требуются все собственные значения и собственные векторы)
- вычисление части спектра (минимальные/максимальные собственные значения, собственные значения в заданной области)





Вычисление собственных значений через характеристическое уравнение

Задача на собственные значения имеет вид

$$Ax = \lambda x$$

или

$$(A - \lambda I)x = 0,$$

следовательно, матрица $A - \lambda I$ имеет нетривиальное ядро и должна быть сингулярной.

Это означает, что определитель

$$p(\lambda) = \det(A - \lambda I) = 0.$$

ullet Уравнение называется **характеристическим уравнением**, а левая часть является полиномом степени n.





Вычисление собственных значений через характеристическое уравнение

Задача на собственные значения имеет вид

$$Ax = \lambda x$$
,

или

$$(A - \lambda I)x = 0,$$

следовательно, матрица $A - \lambda I$ имеет нетривиальное ядро и должна быть сингулярной.

Это означает, что определитель

$$p(\lambda) = \det(A - \lambda I) = 0.$$

- Уравнение называется **характеристическим уравнением**, а левая часть является полиномом степени n.
- Полином n-й степени имеет n комплексных корней!



• Теперь мы вернемся к собственным значениям.





- Теперь мы вернемся к собственным значениям.
- Характеристическое уравнение можно использовать для вычисления собственных значений, что приводит к наивному алгоритму:

$$p(\lambda) = \det(A - \lambda I)$$



- Теперь мы вернемся к собственным значениям.
- Характеристическое уравнение можно использовать для вычисления собственных значений, что приводит к наивному алгоритму:

$$p(\lambda) = \det(A - \lambda I)$$

1. Вычислите коэффициенты полинома



- Теперь мы вернемся к собственным значениям.
- Характеристическое уравнение можно использовать для вычисления собственных значений, что приводит к наивному алгоритму:

$$p(\lambda) = \det(A - \lambda I)$$

- 1. Вычислите коэффициенты полинома
- 2. Вычислите корни



- Теперь мы вернемся к собственным значениям.
- Характеристическое уравнение можно использовать для вычисления собственных значений, что приводит к наивному алгоритму:

$$p(\lambda) = \det(A - \lambda I)$$

- 1. Вычислите коэффициенты полинома
- 2. Вычислите корни



- Теперь мы вернемся к собственным значениям.
- Характеристическое уравнение можно использовать для вычисления собственных значений, что приводит к наивному алгоритму:

$$p(\lambda) = \det(A - \lambda I)$$

- 1. Вычислите коэффициенты полинома
- 2. Вычислите корни

Это хорошая идея?



• Мы часто интересуемся вычислением части спектра, например, наибольших или наименьших собственных значений.





- Мы часто интересуемся вычислением части спектра, например, наибольших или наименьших собственных значений.
- Также интересно отметить, что для эрмитовых матриц $(A=A^*)$ собственные значения всегда действительные (докажите это!).



- Мы часто интересуемся вычислением части спектра, например, наибольших или наименьших собственных значений.
- ullet Также интересно отметить, что для эрмитовых матриц $(A=A^*)$ собственные значения всегда действительные (докажите это!).
- Степенной метод является самым простым методом для вычисления наибольшего по модулю собственного значения.





- Мы часто интересуемся вычислением части спектра, например, наибольших или наименьших собственных значений.
- ullet Также интересно отметить, что для эрмитовых матриц $(A=A^*)$ собственные значения всегда действительные (докажите это!).
- Степенной метод является самым простым методом для вычисления наибольшего по модулю собственного значения.
- Он также является нашим первым примером итерационного метода и метода Крылова.





Упражнение

Докажите, что для эрмитовых матриц $(A=A^*)$ собственные значения всегда действительные.



Упражнение

Докажите, что для эрмитовых матриц ($A=A^st$) собственные значения всегда действительные.

Решение.

1. Пусть λ - собственное значение матрицы A, а $x \neq 0$ - собственный вектор, тогда:

$$Ax = \lambda x$$

$$x^*Ax = \lambda x^*x$$



Упражнение

Докажите, что для эрмитовых матриц $(A=A^*)$ собственные значения всегда действительные.

Решение.

1. Пусть λ - собственное значение матрицы A, а $x \neq 0$ - собственный вектор, тогда:

$$Ax = \lambda x$$
$$x^* Ax = \lambda x^* x$$

2. Так как $A=A^*$, выражение слева - эрмитово (вещественное) число, значит:

$$(x^*Ax)^* = x^*A^*x = x^*Ax$$

 $(\lambda x^*x)^* = \lambda^*x^*x = (x^*Ax)^* = \lambda x^*x$



Упражнение

Докажите, что для эрмитовых матриц $(A=A^*)$ собственные значения всегда действительные.

Решение.

1. Пусть λ - собственное значение матрицы A, а $x \neq 0$ - собственный вектор, тогда:

$$Ax = \lambda x$$
$$x^* A x = \lambda x^* x$$

2. Так как $A=A^*$, выражение слева - эрмитово (вещественное) число, значит:

$$(x^*Ax)^* = x^*A^*x = x^*Ax$$

 $(\lambda x^*x)^* = \lambda^*x^*x = (x^*Ax)^* = \lambda x^*x$

3. Поскольку $x^*x > 0$, отсюда следует:

$$\lambda=\lambda^*\Rightarrow\lambda\in\mathbb{R}$$



1. Задача на собственные значения

$$Ax = \lambda x$$
, $\|x\|_2 = 1$ для стабильности.

может быть переписана как итерационный метод.



1. Задача на собственные значения

$$Ax = \lambda x, \quad \|x\|_2 = 1$$
 для стабильности.

может быть переписана как итерационный метод.

2. Этот метод называется **степенным методом** и находит наибольшее по модулю собственное значение матрицы A.



1. Задача на собственные значения

$$Ax = \lambda x, \quad \|x\|_2 = 1$$
 для стабильности.

может быть переписана как итерационный метод.

- 2. Этот метод называется степенным методом и находит наибольшее по модулю собственное значение матрицы A.
- 3. Степенной метод имеет вид

$$x_{k+1} = Ax_k, \quad x_{k+1} := \frac{x_{k+1}}{\|x_{k+1}\|_2}$$

И

$$x_{k+1} \rightarrow v_1$$

где $Av_1 = \lambda_1 v_1$ и λ_1 является наибольшим собственным значением и v_1 является соответствующим собственным вектором.



1. Задача на собственные значения

$$Ax = \lambda x, \quad \|x\|_2 = 1$$
 для стабильности.

может быть переписана как итерационный метод.

- 2. Этот метод называется степенным методом и находит наибольшее по модулю собственное значение матрицы A.
- 3. Степенной метод имеет вид

$$x_{k+1} = Ax_k, \quad x_{k+1} := \frac{x_{k+1}}{\|x_{k+1}\|_2}$$

и

$$x_{k+1} \to v_1$$

где $Av_1 = \lambda_1 v_1$ и λ_1 является наибольшим собственным значением и v_1 является соответствующим собственным вектором.

4. На (k+1)-й итерации приближение к λ_1 может быть найдено как

$$\lambda^{(k+1)} = (Ax_{k+1}, x_{k+1}),$$



1. Задача на собственные значения

$$Ax = \lambda x, \quad \|x\|_2 = 1$$
 для стабильности.

может быть переписана как итерационный метод.

- 2. Этот метод называется степенным методом и находит наибольшее по модулю собственное значение матрицы A.
- 3. Степенной метод имеет вид

$$x_{k+1} = Ax_k, \quad x_{k+1} := \frac{x_{k+1}}{\|x_{k+1}\|_2}$$

и

$$x_{k+1} \rightarrow v_1$$

где $Av_1 = \lambda_1 v_1$ и λ_1 является наибольшим собственным значением и v_1 является соответствующим собственным вектором.

4. На (k+1)-й итерации приближение к λ_1 может быть найдено как

$$\lambda^{(k+1)} = (Ax_{k+1}, x_{k+1}),$$

5. Заметим, что $\lambda^{(k+1)}$ не требуется для (k+2)-й итерации, но может быть полезно для измерения ошибки на каждой итерации: $||Ax_{k+1} - \lambda^{(k+1)}x_{k+1}||$.



1. Задача на собственные значения

$$Ax = \lambda x$$
, $||x||_2 = 1$ для стабильности.

может быть переписана как итерационный метод.

- 2. Этот метод называется **степенным методом** и находит наибольшее по модулю собственное значение матрицы A.
- 3. Степенной метод имеет вид

$$x_{k+1} = Ax_k, \quad x_{k+1} := \frac{x_{k+1}}{\|x_{k+1}\|_2}$$

И

$$x_{k+1} \to v_1$$

где $Av_1=\lambda_1v_1$ и λ_1 является наибольшим собственным значением и v_1 является соответствующим собственным вектором.

4. На (k+1)-й итерации приближение к λ_1 может быть найдено как

$$\lambda^{(k+1)} = (Ax_{k+1}, x_{k+1}),$$

- 5. Заметим, что $\lambda^{(k+1)}$ не требуется для (k+2)-й итерации, но может быть полезно для измерения ошибки на каждой итерации: $\|Ax_{k+1} \lambda^{(k+1)}x_{k+1}\|$.
- 6. Скорость алгоритма геометрическая (линейная), но показатель сходимости равен q^k , где $q=\left|\frac{\lambda_2}{\lambda_1}\right|<1$, для $\lambda_1>\lambda_2\geq\cdots\geq\lambda_n$ и k число итераций.



1. Задача на собственные значения

$$Ax = \lambda x$$
, $||x||_2 = 1$ для стабильности.

может быть переписана как итерационный метод.

- 2. Этот метод называется **степенным методом** и находит наибольшее по модулю собственное значение матрицы A.
- 3. Степенной метод имеет вид

$$x_{k+1} = Ax_k, \quad x_{k+1} := \frac{x_{k+1}}{\|x_{k+1}\|_2}$$

И

$$x_{k+1} \rightarrow v_1$$

где $Av_1=\lambda_1v_1$ и λ_1 является наибольшим собственным значением и v_1 является соответствующим собственным вектором.

4. На (k+1)-й итерации приближение к λ_1 может быть найдено как

$$\lambda^{(k+1)} = (Ax_{k+1}, x_{k+1}),$$

- 5. Заметим, что $\lambda^{(k+1)}$ не требуется для (k+2)-й итерации, но может быть полезно для измерения ошибки на каждой итерации: $\|Ax_{k+1} \lambda^{(k+1)}x_{k+1}\|$.
- 6. Скорость алгоритма геометрическая (линейная), но показатель сходимости равен q^k , где $q=\left|\frac{\lambda_2}{\lambda_1}\right|<1$, для $\lambda_1>\lambda_2\geq\cdots\geq\lambda_n$ и k число итераций.
- 7. Это означает, что сходимость может быть произвольно малой. Чтобы доказать это, достаточно рассмотреть 2×2 диагональную матрицу.



• Степенной метод дает оценку наибольшего по модулю собственного значения или спектрального радиуса данной матрицы



- Степенной метод дает оценку наибольшего по модулю собственного значения или спектрального радиуса данной матрицы
- ullet Один шаг требует одного матрично-векторного произведения. Если матрица позволяет $\mathcal{O}(n)$ matvec (например, она разрежена), то степенной метод можно использовать для больших n.



- Степенной метод дает оценку наибольшего по модулю собственного значения или спектрального радиуса данной матрицы
- ullet Один шаг требует одного матрично-векторного произведения. Если матрица позволяет $\mathcal{O}(n)$ matvec (например, она разрежена), то степенной метод можно использовать для больших n.
- Сходимость может быть медленной



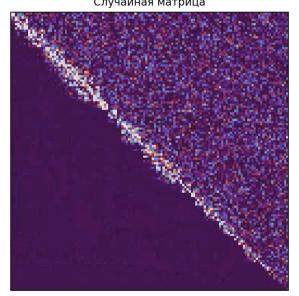
- Степенной метод дает оценку наибольшего по модулю собственного значения или спектрального радиуса данной матрицы
- ullet Один шаг требует одного матрично-векторного произведения. Если матрица позволяет $\mathcal{O}(n)$ matvec (например, она разрежена), то степенной метод можно использовать для больших n.
- Сходимость может быть медленной
- Если нужна только грубая оценка, достаточно нескольких итераций



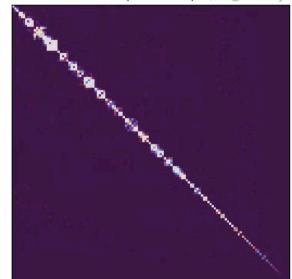
- Степенной метод дает оценку наибольшего по модулю собственного значения или спектрального радиуса данной матрицы
- ullet Один шаг требует одного матрично-векторного произведения. Если матрица позволяет $\mathcal{O}(n)$ matvec (например, она разрежена), то степенной метод можно использовать для больших n.
- Сходимость может быть медленной
- Если нужна только грубая оценка, достаточно нескольких итераций
- Решение вектор находится в подпространстве Крылова $\{x_0, Ax_0, \dots, A^kx_0\}$ и имеет вид μA^kx_0 , где μ является нормирующим коэффициентом.



В следующих сериях: QR-алгоритм Случайная матрица



Симметричная матрица @fminxyz

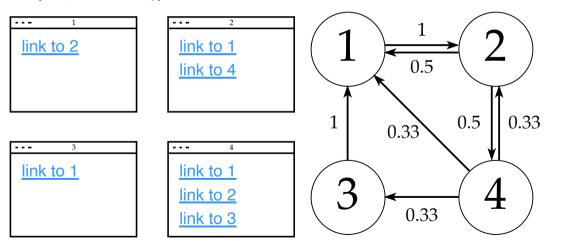


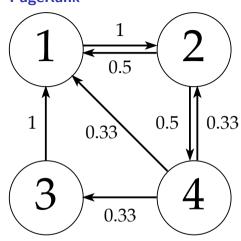


PageRank



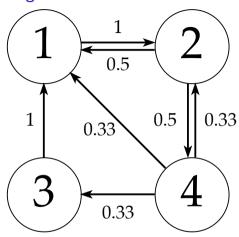
Рассмотрим простой пример. Предположим, что у нас есть 4 веб-сайта с некоторыми ссылками. Наша цель --- понять, насколько важен каждый из этих сайтов. Очевидно, что мы можем переформулировать эту проблему в терминах ориентированных графов. Здесь каждый узел представляет веб-сайт, а каждое ребро описывает ссылку с одного сайта на другой.





Таким образом, мы можем ввести матрицу переходов A:

$$A = \begin{bmatrix} 0 & \frac{1}{2} & 1 & \frac{1}{3} \\ 1 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$



Таким образом, мы можем ввести матрицу переходов A:

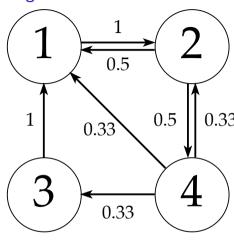
$$A = \begin{bmatrix} 0 & \frac{1}{2} & 1 & \frac{1}{3} \\ 1 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

описывает важность каждого веб-сайта.

Давайте введём вектор PageRank x, который

$$\mathbf{x} = (x_1, x_2, x_3, x_4)^\mathsf{T},$$
 где x_i - важность i -го веб-сайта

PageRank



Таким образом, мы можем ввести матрицу переходов A:

$$A = \begin{bmatrix} 0 & \frac{1}{2} & 1 & \frac{1}{3} \\ 1 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

описывает важность каждого веб-сайта.

Давайте введём вектор PageRank x, который

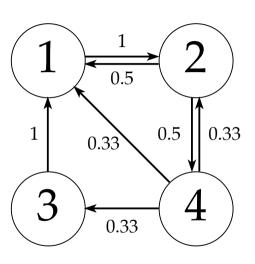
$$\mathbf{x} = (x_1, x_2, x_3, x_4)^\mathsf{T},$$
 где x_i - важность i -го веб-сайта

Предположим, что начальная важность равномерно распределена между всеми узлами. Тогда:

$$\mathbf{x}^0 = (0.25, 0.25, 0.25, 0.25)^\mathsf{T}$$

Каждая входящая ссылка увеличивает важность узла. Таким образом, этот обновление может быть записано как умножение матрицы на вектор:

$$\mathbf{x}^1 = A \cdot \mathbf{x}^0 = (0.46, 0.33, 0.08, 0.125)^\mathsf{T}$$



Повторяя те же операции, мы можем легко увидеть сходимость:

$$\mathbf{x}^2 = A \cdot \mathbf{x}^1 = (0.29, 0.50, 0.04, 0.17)^\mathsf{T}$$

$$\mathbf{x}^3 = A \cdot \mathbf{x}^2 = (0.35, 0.35, 0.06, 0.25)^\mathsf{T}$$

$$\mathbf{x}^{14} = A \cdot \mathbf{x}^{13} = (0.33, 0.40, 0.07, 0.20)^{\mathsf{T}}$$

$$\mathbf{x}^{15} = A \cdot \mathbf{x}^{14} = (0.33, 0.40, 0.07, 0.20)^{\mathsf{T}}$$

Выполните упражнение **P**ageRank.