Сингулярное разложение и главные компоненты

Обобщение диагонализации

Краткий план:

- Жорданова нормальная форма.

Краткий план:

- Жорданова нормальная форма.
- Сингулярное разложение.

Краткий план:

- Жорданова нормальная форма.
- Сингулярное разложение.
- Доказательство существования.

Не все матрицы диагонализуемы

Утверждение

Квадратная матрица A размера $n \times n$ диагонализуема, если у неё найдётся n линейно независимых собственных векторов.

В этом случае A представима в виде $A = PDP^{-1}$,

где P — матрица из собственных векторов, D — диагональная матрица из собственных значений.

Не все матрицы диагонализуемы

Утверждение

Квадратная матрица A размера $n \times n$ диагонализуема, если у неё найдётся n линейно независимых собственных векторов.

В этом случае A представима в виде $A = PDP^{-1}$,

где P — матрица из собственных векторов, D — диагональная матрица из собственных значений.

Утверждение

У симметричной матрицы A размера $n \times n$ найдётся n ортогональных собственных векторов единичной длины.

С их помощью матрица A представима в виде

$$A = PDP^T.$$

А если не везёт?

Что делать, если у матрицы A размера $n \times n$ меньше, чем n независимых собственных векторов?

А если не везёт?

Что делать, если у матрицы A размера $n \times n$ меньше, чем n независимых собственных векторов?

Утверждение

Любая квадратная матрица A представима в виде

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

где жорданова нормальная форма J содержит на диагонали жордановы клетки J_i:

$$J = \begin{pmatrix} J_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & J_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & J_k \end{pmatrix}, \quad J_i = \begin{pmatrix} \lambda_i & 1 & \dots & 0 \\ 0 & \lambda_i & \dots & 0 \\ 0 & \dots & \dots & 1 \\ 0 & 0 & \dots & \lambda_i \end{pmatrix}$$

А если матрица не квадратная?

Симметричные матрицы и матрицы с n независимыми собственными векторами диагонализуемы.

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

А если матрица не квадратная?

Симметричные матрицы и матрицы с n независимыми собственными векторами диагонализуемы.

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

Любая квадратная матрица приводима к жордановой нормальной форме.

$$A = PJP^{-1}$$

А если матрица не квадратная?

Симметричные матрицы и матрицы с n независимыми собственными векторами диагонализуемы.

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

Любая квадратная матрица приводима к жордановой нормальной форме.

$$A = PJP^{-1}$$

Любая матрица имеет сингулярное разложение!

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

Сингулярное разложение

Утверждение

Любой линейный оператор из $\mathbb{R}^k$ в $\mathbb{R}^n$ представляет собой последовательность действий:

Сингулярное разложение

Утверждение

Любой линейный оператор из $\mathbb{R}^k$ в $\mathbb{R}^n$ представляет собой последовательность действий:

1. Ортогональное преобразование из $\mathbb{R}^k$ в $\mathbb{R}^k$. Сохраняет длины векторов и углы между ними.

Сингулярное разложение

Утверждение

Любой линейный оператор из $\mathbb{R}^k$ в $\mathbb{R}^n$ представляет собой последовательность действий:

1. Ортогональное преобразование из $\mathbb{R}^k$ в $\mathbb{R}^k$. Сохраняет длины векторов и углы между ними.
2. Растягивание компонент вектора в $\mathbb{R}^k$.

Сингулярное разложение

Утверждение

Любой линейный оператор из $\mathbb{R}^k$ в $\mathbb{R}^n$ представляет собой последовательность действий:

1. Ортогональное преобразование из $\mathbb{R}^k$ в $\mathbb{R}^k$. Сохраняет длины векторов и углы между ними.
2. Растягивание компонент вектора в $\mathbb{R}^k$.
3. Переход из $\mathbb{R}^k$ в $\mathbb{R}^n$ путём дописывания нулей в вектор, или зачёркивания части координат.

Сингулярное разложение

Утверждение

Любой линейный оператор из $\mathbb{R}^k$ в $\mathbb{R}^n$ представляет собой последовательность действий:

1. Ортогональное преобразование из $\mathbb{R}^k$ в $\mathbb{R}^k$. Сохраняет длины векторов и углы между ними.
2. Растягивание компонент вектора в $\mathbb{R}^k$.
3. Переход из $\mathbb{R}^k$ в $\mathbb{R}^n$ путём дописывания нулей в вектор, или зачёркивания части координат.
4. Ортогональное преобразование из $\mathbb{R}^n$ в $\mathbb{R}^n$. Сохраняет длины векторов и углы между ними.

Сингулярное разложение

Утверждение

Любой линейный оператор из $\mathbb{R}^k$ в $\mathbb{R}^n$ представляет собой последовательность действий:

1. Ортогональное преобразование из $\mathbb{R}^k$ в $\mathbb{R}^k$. Сохраняет длины векторов и углы между ними.
2. Растягивание компонент вектора в $\mathbb{R}^k$.
3. Переход из $\mathbb{R}^k$ в $\mathbb{R}^n$ путём дописывания нулей в вектор, или зачёркивания части координат.
4. Ортогональное преобразование из $\mathbb{R}^n$ в $\mathbb{R}^n$. Сохраняет длины векторов и углы между ними.

Все эти действия мы рассмотрели на первой лекции!

Сингулярное разложение проекции

Оператор H $: \mathbb{R}^n \to \mathbb{R}^n$ проецирует векторы на линейную оболочку M.

Сингулярное разложение проекции

Оператор H $: \mathbb{R}^n \to \mathbb{R}^n$ проецирует векторы на линейную оболочку M.

Выберем ортогональный базис $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ в M.

Сингулярное разложение проекции

Оператор H $: \mathbb{R}^n \to \mathbb{R}^n$ проецирует векторы на линейную оболочку M.

Выберем ортогональный базис $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ в M.

Дополним его до ортогонального базиса в $\mathbb{R}^n$ векторами $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$.

Сингулярное разложение проекции

Оператор H $: \mathbb{R}^n \to \mathbb{R}^n$ проецирует векторы на линейную оболочку M.

Выберем ортогональный базис $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ в M.

Дополним его до ортогонального базиса в $\mathbb{R}^n$ векторами $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$.

1. Повернём-отразим пространство, чтобы $\mathbf{v}_1, \ldots, \mathbf{v}_n$ перешли в $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Сингулярное разложение проекции

Оператор H : $\mathbb{R}^n \to \mathbb{R}^n$ проецирует векторы на линейную оболочку M.

Выберем ортогональный базис $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ в M.

Дополним его до ортогонального базиса в $\mathbb{R}^n$ векторами $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$.

1. Повернём-отразим пространство, чтобы $\mathbf{v}_1, \ldots, \mathbf{v}_n$ перешли в $\mathbf{e}_1, \ldots, \mathbf{e}_n$.
2. Домножим первые k компонент вектора на 1, а остальные на 0.

Сингулярное разложение проекции

Оператор H $: \mathbb{R}^n \to \mathbb{R}^n$ проецирует векторы на линейную оболочку M.

Выберем ортогональный базис $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ в M.

Дополним его до ортогонального базиса в $\mathbb{R}^n$ векторами $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$.

1. Повернём-отразим пространство, чтобы $\mathbf{v}_1, \ldots, \mathbf{v}_n$ перешли в $\mathbf{e}_1, \ldots, \mathbf{e}_n$.
2. Домножим первые k компонент вектора на 1, а остальные на 0.
3. Смены размерности нет.

Сингулярное разложение проекции

Оператор H : $\mathbb{R}^n \to \mathbb{R}^n$ проецирует векторы на линейную оболочку M.

Выберем ортогональный базис $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ в M.

Дополним его до ортогонального базиса в $\mathbb{R}^n$ векторами $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$.

1. Повернём-отразим пространство, чтобы $\mathbf{v}_1, \ldots, \mathbf{v}_n$ перешли в $\mathbf{e}_1, \ldots, \mathbf{e}_n$.
2. Домножим первые k компонент вектора на 1, а остальные на 0.
3. Смены размерности нет.
4. Повернём-отразим пространство, чтобы $\mathbf{e}_1, \ldots, \mathbf{e}_n$ перешли в $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Сингулярное разложение

Утверждение

Любую матрицу A размера $n \times k$ можно представить в виде

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

где матрица U размера $n \times n$ — ортогональная, $U^T U = \mathrm{I}$,

матрица V размера $k \times k$ — ортогональная, $V^T V = \mathrm{I}$,

матрица Σ размера $n \times k$ — диагональная.

Сингулярное разложение

Утверждение

Любую матрицу A размера $n \times k$ можно представить в виде

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

где матрица U размера $n \times n$ — ортогональная, $U^T U = \mathrm{I}$,

матрица V размера $k \times k$ — ортогональная, $V^T V = \mathrm{I}$,

матрица Σ размера $n \times k$ — диагональная.

Данное разложение также называется SVD-разложением, singular value decomposition.

Присмотримся к матрицам

Если $n \geq k$, то SVD-разложение примет вид

$$A = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \ldots & \mathbf{u}_n \\ | & & | \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma_k \\ 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_k^T & - \end{pmatrix}$$

Присмотримся к матрицам

Если $n \geq k$, то SVD-разложение примет вид

$$A = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_n \\ | & & | \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_k \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_k^T & - \end{pmatrix}$$

Числа σ_i называются сингулярными числами матрицы A.

Для удобства их сортируют по убыванию $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots$.

Зачем нужно SVD-разложение?

Сильно упрощает многие вычисления.

Зачем нужно SVD-разложение?

Сильно упрощает многие вычисления.

Показывает внутренний мир матрицы.

Зачем нужно SVD-разложение?

Сильно упрощает многие вычисления.

Показывает внутренний мир матрицы.

Существует быстрая и устойчивая итеративная процедура нахождения SVD-разложения.

Раскладываем все матрицы!

Сингулярное разложение для A^T:

$$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$$

Раскладываем все матрицы!

Сингулярное разложение для A^T:

$$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$$

Сингулярное разложение для $A^T A$:

$$A^T A = V\Sigma^T U^T \cdot U\Sigma V^T = V\Sigma^T\Sigma V^T$$

Раскладываем все матрицы!

Сингулярное разложение для A^T:

$$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$$

Сингулярное разложение для $A^T A$:

$$A^T A = V\Sigma^T U^T \cdot U\Sigma V^T = V\Sigma^T \Sigma V^T$$

Сингулярное разложение для AA^T:

$$AA^T = U\Sigma V^T \cdot V\Sigma^T U^T = U\Sigma\Sigma^T U^T$$

Существование разложения

Наша задача предъявить разложение $A = U\Sigma V^T$.

Для удобства будем считать, что $n \geq k$.

Существование разложения

Наша задача предъявить разложение $A = U\Sigma V^T$.

Для удобства будем считать, что $n \geq k$.

Доказательство

1. Матрица $A^T A$ является матрицей Грама. Она положительно определена и симметрична. А потому представима в виде $A^T A = VDV^T$.

Существование разложения

Наша задача предъявить разложение $A = U\Sigma V^T$.

Для удобства будем считать, что $n \geq k$.

Доказательство

1. Матрица $A^T A$ является матрицей Грама. Она положительно определена и симметрична. А потому представима в виде $A^T A = VDV^T$.
2. Диагональные элементы D неотрицательны. Поэтому D представима в виде $\Sigma^T \Sigma$.

Существование разложения

Наша задача предъявить разложение $A = U\Sigma V^T$.

Для удобства будем считать, что $n \geq k$.

Доказательство

1. Матрица $A^T A$ является матрицей Грама. Она положительно определена и симметрична. А потому представима в виде $A^T A = V D V^T$.
2. Диагональные элементы D неотрицательны. Поэтому D представима в виде $\Sigma^T \Sigma$.
3. Осталось найти U из целевого разложения:

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

Существование разложения

Наша задача предъявить разложение $A = U\Sigma V^T$.

Для удобства будем считать, что $n \geq k$.

Доказательство

1. Матрица $A^T A$ является матрицей Грама. Она положительно определена и симметрична. А потому представима в виде $A^T A = V D V^T$.
2. Диагональные элементы D неотрицательны. Поэтому D представима в виде $\Sigma^T \Sigma$.
3. Осталось найти U из целевого разложения:

$$A = U\Sigma V^T \text{ или } AV = U\Sigma$$

Существование разложения

Уже нашли Σ и V. Осталось найти U из $AV = U\Sigma$.

Существование разложения

Уже нашли Σ и V. Осталось найти U из $AV = U\Sigma$.

Окончание доказательства

4. Находим вектора $\mathbf{u}_i$ по очереди:

$$\mathbf{u}_1 = A\mathbf{v}_1/\sigma_1,$$

Существование разложения

Уже нашли Σ и V. Осталось найти U из $AV = U\Sigma$.

Окончание доказательства

4. Находим вектора $\mathbf{u}_i$ по очереди:

$$\mathbf{u}_1 = A\mathbf{v}_1/\sigma_1,\ \mathbf{u}_2 = A\mathbf{v}_2/\sigma_2, \ldots$$

Существование разложения

Уже нашли Σ и V. Осталось найти U из $AV = U\Sigma$.

Окончание доказательства

4. Находим вектора $\mathbf{u}_i$ по очереди:

$$\mathbf{u}_1 = A\mathbf{v}_1/\sigma_1,\ \mathbf{u}_2 = A\mathbf{v}_2/\sigma_2, \ldots$$

5. Вектора $\mathbf{v}_i$ кончатся раньше $\mathbf{u}_i$.

Существование разложения

Уже нашли Σ и V. Осталось найти U из $AV = U\Sigma$.

Окончание доказательства

4. Находим вектора $\mathbf{u}_i$ по очереди:

$$\mathbf{u}_1 = A\mathbf{v}_1/\sigma_1,\ \mathbf{u}_2 = A\mathbf{v}_2/\sigma_2, \ldots$$

5. Вектора $\mathbf{v}_i$ кончатся раньше $\mathbf{u}_i$. Оставшиеся $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$ выберем произвольными, чтобы U была ортогональной матрицей.

Авторы современного алгоритма

Джин Говард Голуб

Уильям Мортон Кэхэн

Отец плавающей точки

wikipedia.org / П. Биркен, Д. Биркен, фото номера: П.М. Кроненберг, Клив Молер

Поиск SVD разложения

Это видеофрагмент с доской, слайдов здесь нет :)

Нахождение проекции при известном SVD

Это видеофрагмент с доской, слайдов здесь нет :)

Немного статистики

Краткий план:

- Выборочная дисперсия.

Краткий план:

- Выборочная дисперсия.
- Стандартизация.

Краткий план:

- Выборочная дисперсия.
- Стандартизация.
- Выборочная корреляция.

Центрирование

Определение

Для вектора чисел $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ выборочным средним называют величину

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Центрирование

Определение

Для вектора чисел $\mathbf{x} = (x_1, x_2, \dots, x_n)$ выборочным средним называют величину

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Определение

Центрирование переменной — переход от набора чисел $(x_1, x_2, \dots, x_n)$ к набору чисел $(x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$.

Центрирование

Определение

Для вектора чисел $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ выборочным средним называют величину

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Определение

Центрирование переменной — переход от набора чисел $(x_1, x_2, \ldots, x_n)$ к набору чисел $(x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})$.

Пример. $(4, 2, 3, 3) \to (1, -1, 0, 0)$.

Свойства центрирования

Утверждение

Если $\mathbf{x}'$ — это центрированный вектор $\mathbf{x}$, $y_i = x_i - \bar{x}$, то $\bar{x}' = 0$.

Свойства центрирования

Утверждение

Если $\mathbf{x}'$ — это центрированный вектор $\mathbf{x}$, $y_i = x_i - \bar{x}$, то $\bar{x}' = 0$.

$$\sum_{i=1}^{n} x_i' = \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = 0$$

Свойства центрирования

Утверждение

Если $\mathbf{x}'$ — это центрированный вектор $\mathbf{x}$, $y_i = x_i - \bar{x}$, то $\bar{x}' = 0$.

$$\sum_{i=1}^{n} x_i' = \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = 0$$

С геометрической точки зрения:

$(\bar{x}, \bar{x}, \dots, \bar{x})$ — проекция вектора $\mathbf{x}$ на Span $\mathbf{v}$, где $\mathbf{v} = (1, 1, \dots, 1)$.

Свойства центрирования

Утверждение

Если $\mathbf{x}'$ — это центрированный вектор $\mathbf{x}$, $y_i = x_i - \bar{x}$, то $\bar{x}' = 0$.

$$\sum_{i=1}^{n} x_i' = \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = 0$$

С геометрической точки зрения:

$(\bar{x}, \bar{x}, \ldots, \bar{x})$ — проекция вектора $\mathbf{x}$ на Span $\mathbf{v}$, где $\mathbf{v} = (1, 1, \ldots, 1)$.

$(x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})$ — проекция вектора $\mathbf{x}$ на Span$^\perp$ $\mathbf{v}$.

Выборочная дисперсия

Определение

Выборочной дисперсией набора чисел $(x_1, x_2, \ldots, x_n)$ называют величину $\|\mathbf{x}'\|^2/(n-1)$, где вектор $\mathbf{x}'$ — это центрированный вектор $\mathbf{x}$.

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

Выборочная дисперсия

Определение

Выборочной дисперсией набора чисел $(x_1, x_2, \ldots, x_n)$ называют величину $\|\mathbf{x}'\|^2/(n-1)$, где вектор $\mathbf{x}'$ — это центрированный вектор $\mathbf{x}$.

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

Выборочная дисперсия вектора измеряет «разброс» x_i, насколько далеки x_i от своего среднего $\bar{x}$.

Стандартное отклонение

Если x_i измеряется в сантиментрах, то выборочная дисперсия измеряется в квадратных сантиметрах.

Стандартное отклонение

Если x_i измеряется в сантиментрах, то выборочная дисперсия измеряется в квадратных сантиметрах.

Определение

Выборочным стандартным отклонением набора $\mathbf{x}$ называется корень из выборочной дисперсии

$$sd(\mathbf{x}) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

Стандартное отклонение

Если x_i измеряется в сантиментрах, то выборочная дисперсия измеряется в квадратных сантиметрах.

Определение

Выборочным стандартным отклонением набора $\mathbf{x}$ называется корень из выборочной дисперсии

$$sd(\mathbf{x}) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

Стандартное отклонение

Если x_i измеряется в сантиментрах, то выборочная дисперсия измеряется в квадратных сантиметрах.

Определение

Выборочным стандартным отклонением набора $\mathbf{x}$ называется корень из выборочной дисперсии

$$sd(\mathbf{x}) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

Если x_i измеряется в сантиментрах, то выборочное стандартное отклонение измеряется в сантиметрах.

Стандартизация

Популярным вариантом масштабирования переменной является переход к безразмерной величине по правилу:

$$x_i \rightarrow x_i' = \frac{x_i - \bar{x}}{sd(\mathbf{x})}.$$

Стандартизация

Популярным вариантом масштабирования переменной является переход к безразмерной величине по правилу:

$$x_i \to x_i' = \frac{x_i - \bar{x}}{sd(\mathbf{x})}.$$

После стандартизации величина имеет нулевое среднее $\bar{x}' = 0$ и единичное стандартное отлонение $sd(\mathbf{x}') = 1$.

Выборочная корреляция

Определение

Выборочной корреляцией двух наборов чисел $\mathbf{x}$ и $\mathbf{y}$ называют косинус угла между их центрированными версиями:

$$\rho(\mathbf{x}, \mathbf{y}) = \cos \angle(\mathbf{x}', \mathbf{y}') = \frac{\langle \mathbf{x}', \mathbf{y}' \rangle}{\|\mathbf{x}'\| \|\mathbf{y}'\|},$$

где $x_i' = x_i - \bar{x}$ и $y_i' = y_i - \bar{y}$.

Выборочная корреляция

Определение

Выборочной корреляцией двух наборов чисел $\mathbf{x}$ и $\mathbf{y}$ называют косинус угла между их центрированными версиями:

$$\rho(\mathbf{x}, \mathbf{y}) = \cos \angle(\mathbf{x}', \mathbf{y}') = \frac{\langle \mathbf{x}', \mathbf{y}' \rangle}{\|\mathbf{x}'\| \|\mathbf{y}'\|},$$

где $x_i' = x_i - \bar{x}$ и $y_i' = y_i - \bar{y}$.

Не определена, если $\mathbf{x}$ или $\mathbf{y}$ состоит из одинаковых чисел.

Выборочная корреляция

Определение

Выборочной корреляцией двух наборов чисел $\mathbf{x}$ и $\mathbf{y}$ называют косинус угла между их центрированными версиями:

$$\rho(\mathbf{x}, \mathbf{y}) = \cos \angle(\mathbf{x}', \mathbf{y}') = \frac{\langle \mathbf{x}', \mathbf{y}' \rangle}{\|\mathbf{x}'\| \|\mathbf{y}'\|},$$

где $x_i' = x_i - \bar{x}$ и $y_i' = y_i - \bar{y}$.

Не определена, если $\mathbf{x}$ или $\mathbf{y}$ состоит из одинаковых чисел.

Лежит в диапазоне от -1 до 1.

Выборочная корреляция

Утверждение

Выборочная корреляция равна

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}.$$

Если после центрирования векторы ортогональны, то выборочная корреляция равна 0.

Выборочная корреляция

Утверждение

Выборочная корреляция равна

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}.$$

Если после центрирования векторы ортогональны, то выборочная корреляция равна 0.

Не чувствительна к масштабированию.

Выборочная корреляционная матрица

Определение

Выборочной корреляционной матрицей переменных $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ называется матрица C, элементами которой являются выборочные корреляции,

$$c_{ij} = \rho(\mathbf{x}_i, \mathbf{x}_j)$$

Выборочная корреляционная матрица

Определение

Выборочной корреляционной матрицей переменных $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ называется матрица C, элементами которой являются выборочные корреляции,

$$c_{ij} = \rho(\mathbf{x}_i, \mathbf{x}_j)$$

Диагональные элементы равны единице, $c_{ii} = 1$.

Выборочная корреляционная матрица

Определение

Выборочной корреляционной матрицей переменных $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ называется матрица C, элементами которой являются выборочные корреляции,

$$c_{ij} = \rho(\mathbf{x}_i, \mathbf{x}_j)$$

Диагональные элементы равны единице, $c_{ii} = 1$.

Если $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ стандартизированы, то $C = X^T X$.

Выборочная корреляционная матрица

Определение

Выборочной корреляционной матрицей переменных $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ называется матрица C, элементами которой являются выборочные корреляции,

$$c_{ij} = \rho(\mathbf{x}_i, \mathbf{x}_j)$$

Диагональные элементы равны единице, $c_{ii} = 1$.

Если $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ стандартизированы, то $C = X^T X$.

Матрица C симметрична и положительно полуопределена.

PCA: максимизация разброса

Краткий план:

- Максимизация выборочной дисперсии.

Краткий план:

- Максимизация выборочной дисперсии.
- Свойства главных компонент.

Метод главных компонент

Есть матрица X исходных наблюдений:

наблюдения отложены по строкам,

а переменные — по столбцам.

Метод главных компонент

Есть матрица X исходных наблюдений:

наблюдения отложены по строкам,

а переменные — по столбцам.

Переменных очень много.

Метод главных компонент

Есть матрица X исходных наблюдений:

наблюдения отложены по строкам,

а переменные — по столбцам.

Переменных очень много.

Хотим иметь небольшое количество переменных, которые бы почти без потерь содержали всю информацию об исходных переменных.

Метод главных компонент

Все исходные переменные предварительно стандартизируем!

Метод главных компонент

Все исходные переменные предварительно стандартизируем!

Для каждого столбца $\mathbf{x}$ выполнены условия $\bar{x} = 0$, $sd(\mathbf{x}) = 1$.

Метод главных компонент

Все исходные переменные предварительно стандартизируем!

Для каждого столбца $\mathbf{x}$ выполнены условия $\bar{x} = 0$, $sd(\mathbf{x}) = 1$.

На базе столбцов $\mathbf{x}_1$, $\mathbf{x}_2$, …, $\mathbf{x}_k$ матрицы X мы создадим $d \leq k$ новых переменных $\mathbf{p}_1$, $\mathbf{p}_2$, …, $\mathbf{p}_d$.

Метод главных компонент

Все исходные переменные предварительно стандартизируем!

Для каждого столбца $\mathbf{x}$ выполнены условия $\bar{x} = 0$, $sd(\mathbf{x}) = 1$.

На базе столбцов $\mathbf{x}_1$, $\mathbf{x}_2$, …, $\mathbf{x}_k$ матрицы X мы создадим $d \leq k$ новых переменных $\mathbf{p}_1$, $\mathbf{p}_2$, …, $\mathbf{p}_d$.

Новые переменные будем создавать по-очереди.

Метод главных компонент

Все исходные переменные предварительно стандартизируем!

Для каждого столбца $\mathbf{x}$ выполнены условия $\bar{x} = 0$, $sd(\mathbf{x}) = 1$.

На базе столбцов $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ матрицы X мы создадим $d \leq k$ новых переменных $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$.

Новые переменные будем создавать по-очереди.

Новые переменные будем называть главными компонентами.

PCA — principal component analysis.

Максимизация разброса

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Максимизация разброса

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Алгоритм

1. Компоненту $\mathbf{p}_1 = X\mathbf{v}_1$ подберём так, чтобы выборочная дисперсия $\mathbf{p}_1$ была максимальной при условии, что $\|\mathbf{v}_1\| = 1$.

Максимизация разброса

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Алгоритм

1. Компоненту $\mathbf{p}_1 = X\mathbf{v}_1$ подберём так, чтобы выборочная дисперсия $\mathbf{p}_1$ была максимальной при условии, что $\|\mathbf{v}_1\| = 1$.
2. Компоненту $\mathbf{p}_2 = X\mathbf{v}_2$ подберём так, чтобы выборочная дисперсия $\mathbf{p}_2$ была максимальной при условии, что $\mathbf{v}_2 \perp \mathbf{v}_1$ и $\|\mathbf{v}_2\| = 1$.

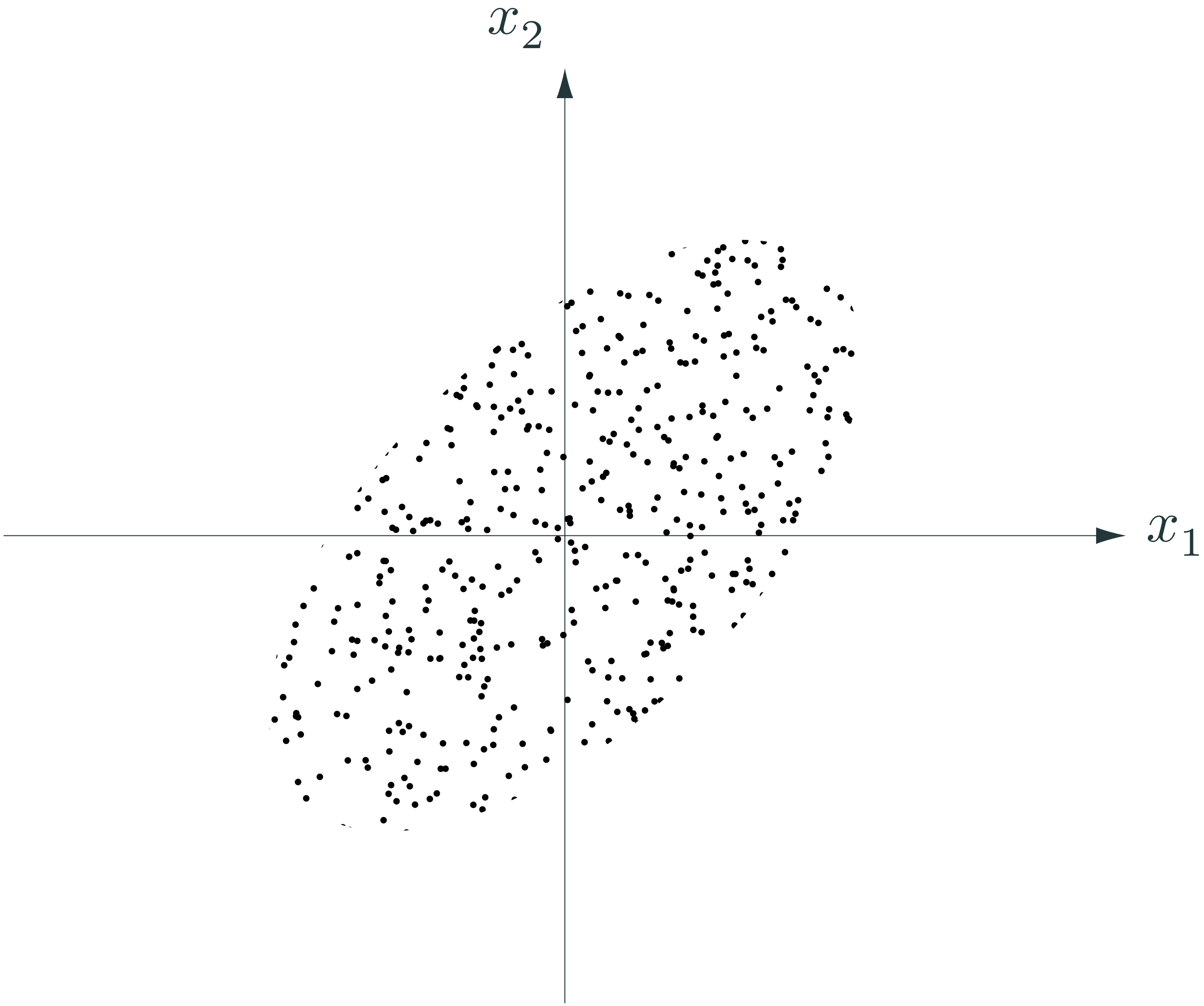
Максимизация разброса

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

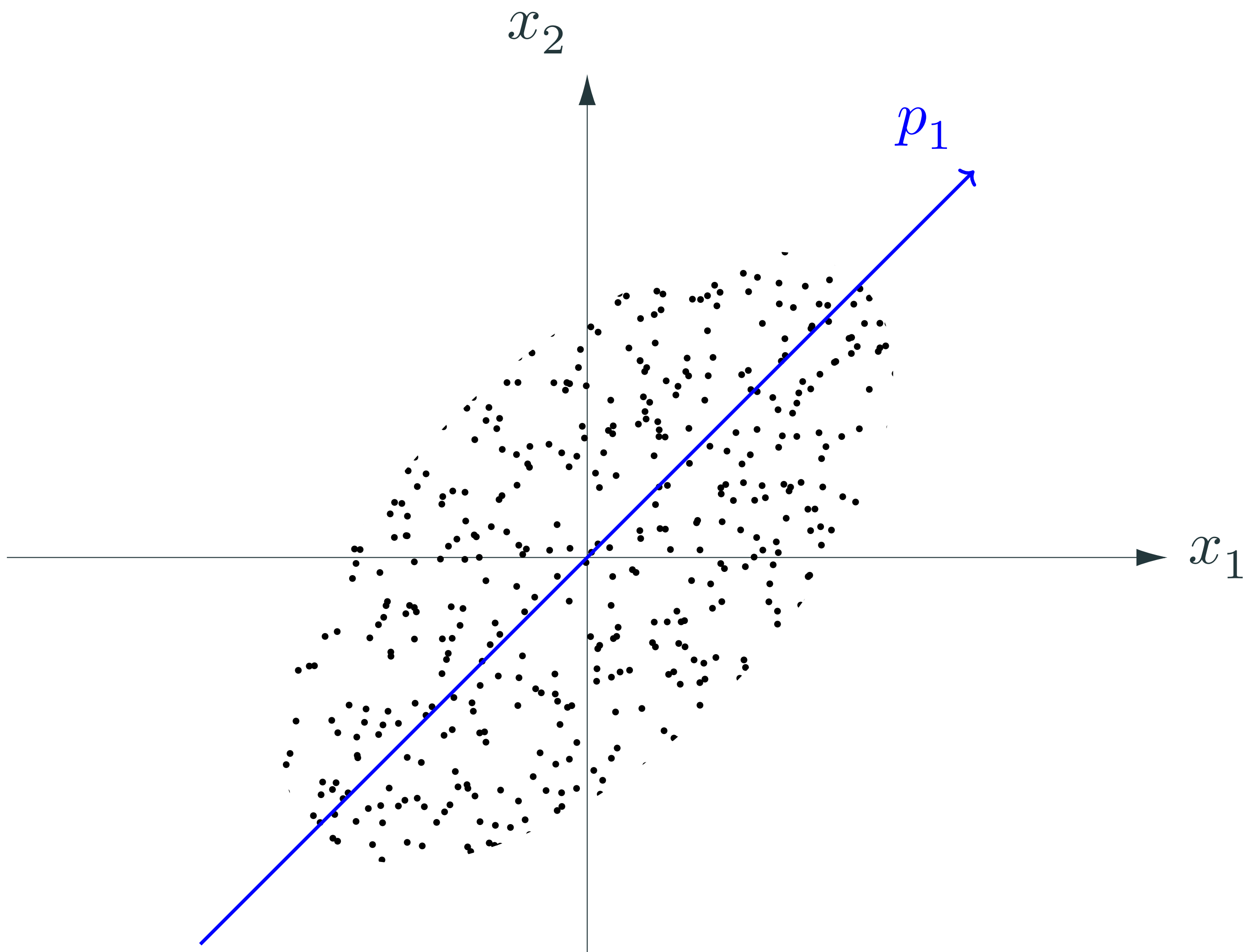
Алгоритм

1. Компоненту $\mathbf{p}_1 = X\mathbf{v}_1$ подберём так, чтобы выборочная дисперсия $\mathbf{p}_1$ была максимальной при условии, что $\|\mathbf{v}_1\| = 1$.
2. Компоненту $\mathbf{p}_2 = X\mathbf{v}_2$ подберём так, чтобы выборочная дисперсия $\mathbf{p}_2$ была максимальной при условии, что $\mathbf{v}_2 \perp \mathbf{v}_1$ и $\|\mathbf{v}_2\| = 1$.
3. Компоненту $\mathbf{p}_3 = X\mathbf{v}_3$ подберём так, чтобы выборочная дисперсия $\mathbf{p}_3$ была максимальной при условии, что $\mathbf{v}_3 \perp \mathbf{v}_2$, $\mathbf{v}_3 \perp \mathbf{v}_1$ и $\|\mathbf{v}_3\| = 1$.
4. …

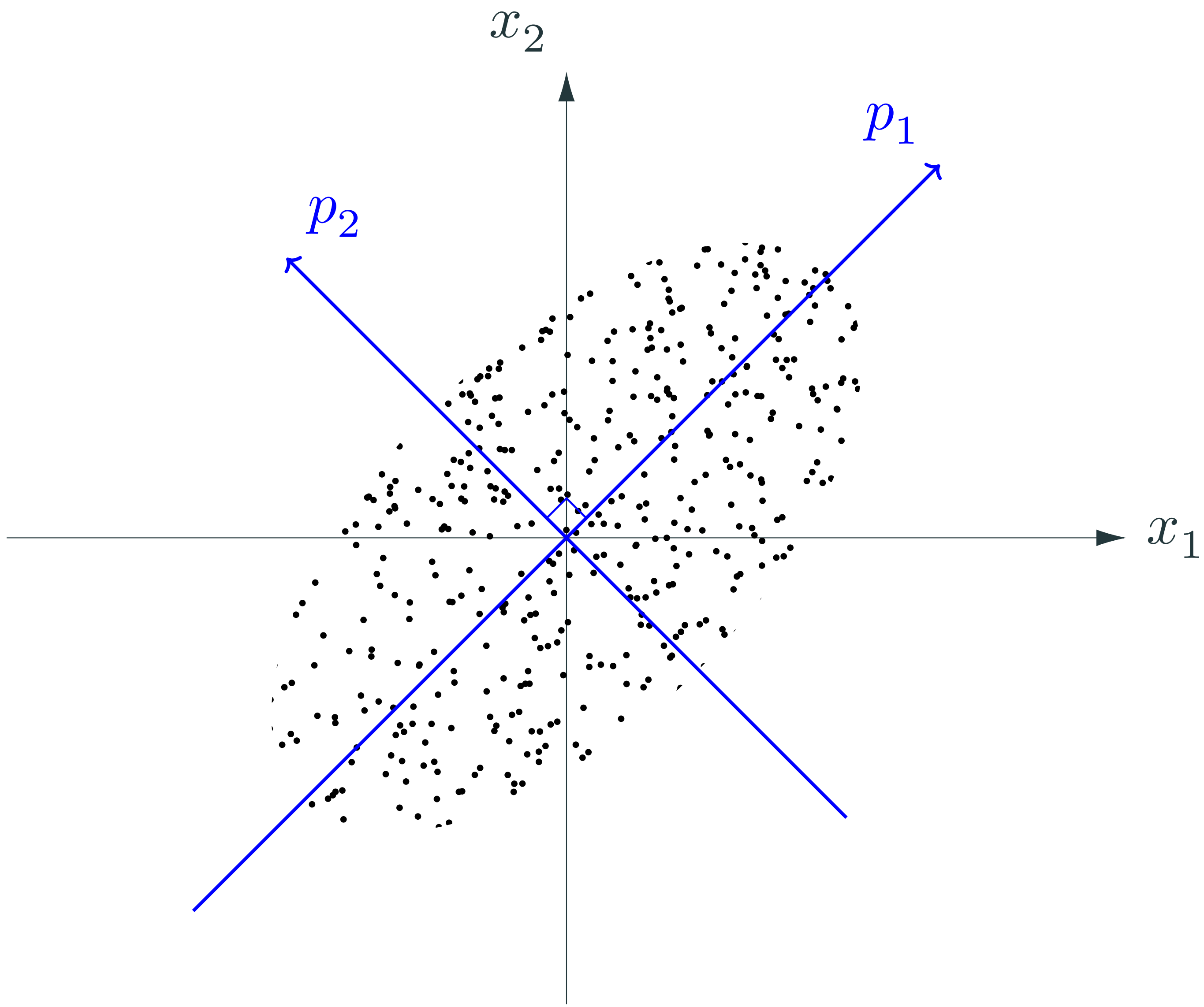
Картинка



Картинка



Картинка



Это завуалированный SVD!

Если $X = U\Sigma V^T$, то $P = XV = U\Sigma$.

Это завуалированный SVD!

Если $X = U\Sigma V^T$, то $P = XV = U\Sigma$.

$$\begin{pmatrix} | & & | \\ \mathbf{p}_1 & \dots & \mathbf{p}_k \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{x}_1 & \dots & \mathbf{x}_k \\ | & & | \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \dots & \mathbf{v}_k \\ | & & | \end{pmatrix}$$

$$\begin{pmatrix} | & & | \\ \mathbf{p}_1 & \dots & \mathbf{p}_k \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_n \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_k \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

Это завуалированный SVD!

Если $X = U\Sigma V^T$, то $P = XV = U\Sigma$.

$$\begin{pmatrix} | & & | \\ \mathbf{p}_1 & \ldots & \mathbf{p}_k \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{x}_1 & \ldots & \mathbf{x}_k \\ | & & | \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \ldots & \mathbf{v}_k \\ | & & | \end{pmatrix}$$

$$\begin{pmatrix} | & & | \\ \mathbf{p}_1 & \ldots & \mathbf{p}_k \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \ldots & \mathbf{u}_n \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma_k \\ 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 \end{pmatrix}$$

Свойства

Главные компоненты находятся из SVD разложения,
$\mathbf{p}_i = \sigma_i \mathbf{u}_i$.

Свойства

Главные компоненты находятся из SVD разложения, $\mathbf{p}_i = \sigma_i \mathbf{u}_i$.

Главные компоненты ортогональны.

Величины σ_i^2 равны квадратам длин $\|\mathbf{p}_i\|^2$ и пропорциональны выборочным дисперсиям $\mathbf{p}_i$.

Значение максимума суммы

$$\|\mathbf{p}_1\|^2 + \ldots + \|\mathbf{p}_d\|^2 = \sigma_1^2 + \ldots + \sigma_d^2$$

одинаково при пошаговой и одновременной максимизации.

Связь с корреляционной матрицей

Если $X = U\Sigma V^T$, то корреляционная матрица имеет вид $C = X^T X = V\Sigma^T \Sigma V^T$.

Связь с корреляционной матрицей

Если $X = U\Sigma V^T$, то корреляционная матрица имеет вид $C = X^T X = V\Sigma^T \Sigma V^T$.

Векторы весов $\mathbf{v}_i$, с которыми исходные переменные входят в компоненты, являются собственными векторами корреляционной матрицы C.

Связь с корреляционной матрицей

Если $X = U\Sigma V^T$, то корреляционная матрица имеет вид $C = X^T X = V\Sigma^T \Sigma V^T$.

Векторы весов $\mathbf{v}_i$, с которыми исходные переменные входят в компоненты, являются собственными векторами корреляционной матрицы C.

Сингулярные значения матрицы X в квадрате являются собственными числами корреляционной матрицы C, $\lambda_i = \sigma_i^2$.

PCA: минимизация ошибки приближения

Краткий план:

- Наилучшая аппроксимация.

Краткий план:

- Наилучшая аппроксимация.
- Минимизация ошибки приближения.

Краткий план:

- Наилучшая аппроксимация.
- Минимизация ошибки приближения.
- Свойства главных компонент.

Метод главных компонент

Есть матрица X исходных наблюдений:

наблюдения отложены по строкам,

а переменные — по столбцам.

Переменных очень много.

Хотим иметь небольшое количество переменных, которые бы почти без потерь содержали всю информацию об исходных переменных.

Поиск наилучшего приближения

Хотим найти матрицу $\hat{X}$ такую, чтобы при заданном ранге rank $\hat{X} = d$ матрица $\hat{X}$ была бы поближе к X:

$$\sum_{ij}(x_{ij} - \hat{x}_{ij})^2 = \mathrm{tr}((X - \hat{X})^T(X - \hat{X})) \to \min$$

Поиск наилучшего приближения

Хотим найти матрицу $\hat{X}$ такую, чтобы при заданном ранге rank $\hat{X} = d$ матрица $\hat{X}$ была бы поближе к X:

$$\sum_{ij}(x_{ij} - \hat{x}_{ij})^2 = \mathrm{tr}((X - \hat{X})^T(X - \hat{X})) \to \min$$

Аппроксимацию с рангом d обозначим $\hat{X}_d$.

Метод главных компонент

Все исходные переменные предварительно стандартизируем!

Для каждого столбца $\mathbf{x}$ выполнены условия $\bar{x} = 0$, $sd(\mathbf{x}) = 1$.

На базе столбцов $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ матрицы X мы создадим $d \leq k$ новых переменных $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$.

Новые переменные будем создавать по-очереди.

Новые переменные будем называть главными компонентами.

Минимизация ошибки приближения

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Минимизация ошибки приближения

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Матрица $\hat{X}_i$ содержит проекции столбцов матрицы X на Span$\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_i\}$.

Алгоритм

1. Компоненту $\mathbf{p}_1 = X\mathbf{v}_1$ подберём так, чтобы матрица $\hat{X}_1$ была наилучшей аппроксимацией X, $\sum_{ij}(x_{ij} - \hat{x}_{ij})^2 \to$ min.

Минимизация ошибки приближения

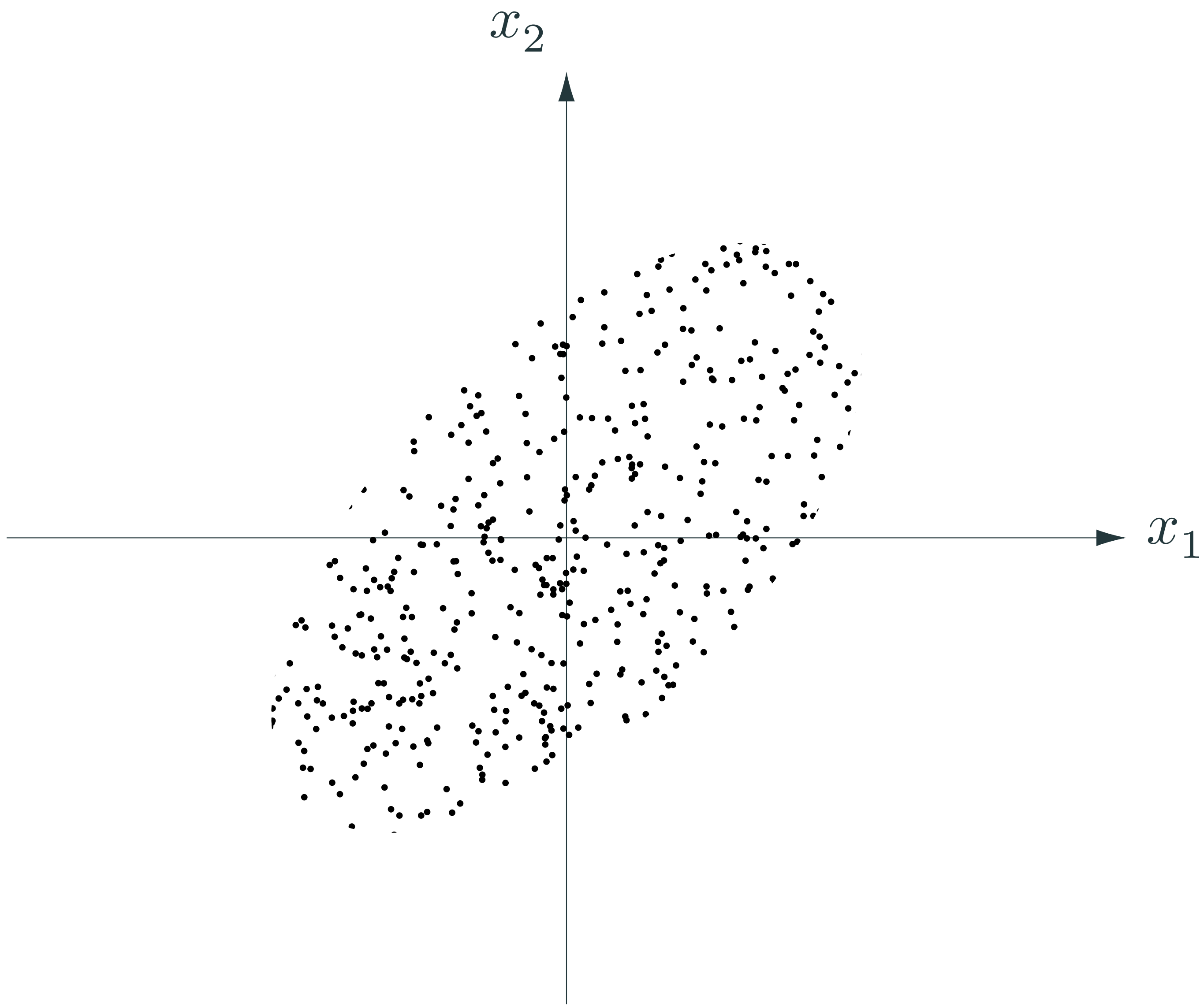
Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Матрица $\hat{X}_i$ содержит проекции столбцов матрицы X на Span$\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_i\}$.

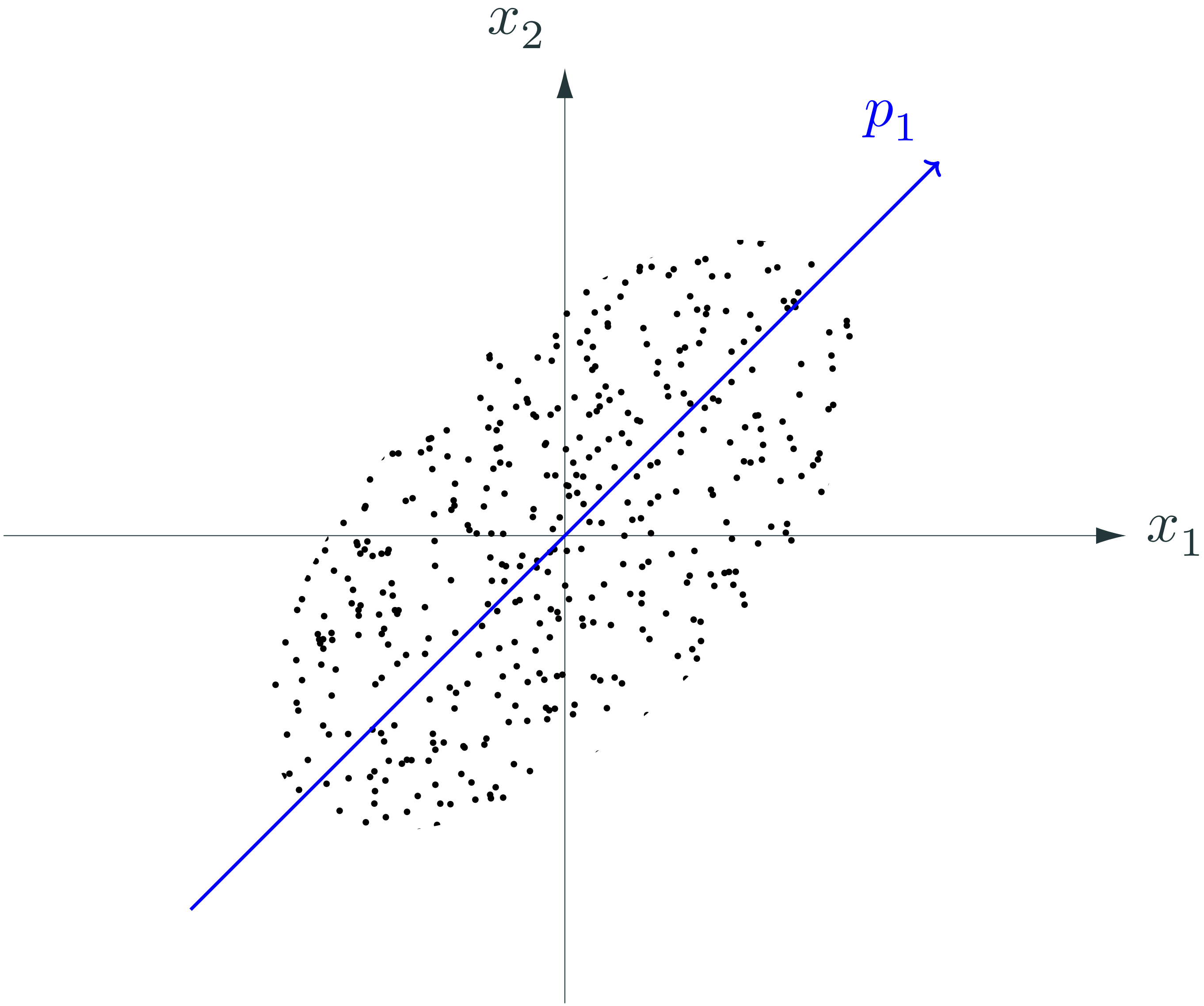
Алгоритм

1. Компоненту $\mathbf{p}_1 = X\mathbf{v}_1$ подберём так, чтобы матрица $\hat{X}_1$ была наилучшей аппроксимацией X, $\sum_{ij}(x_{ij} - \hat{x}_{ij})^2 \to$ min.
2. Компоненту $\mathbf{p}_2 = X\mathbf{v}_2$ подберём так, чтобы матрица $\hat{X}_2$ была наилучшей аппроксимацией X, $\sum_{ij}(x_{ij} - \hat{x}_{ij})^2 \to$ min при условии, что $\mathbf{v}_2 \perp \mathbf{v}_1$ и $\|\mathbf{v}_2\| = 1$.
3. …

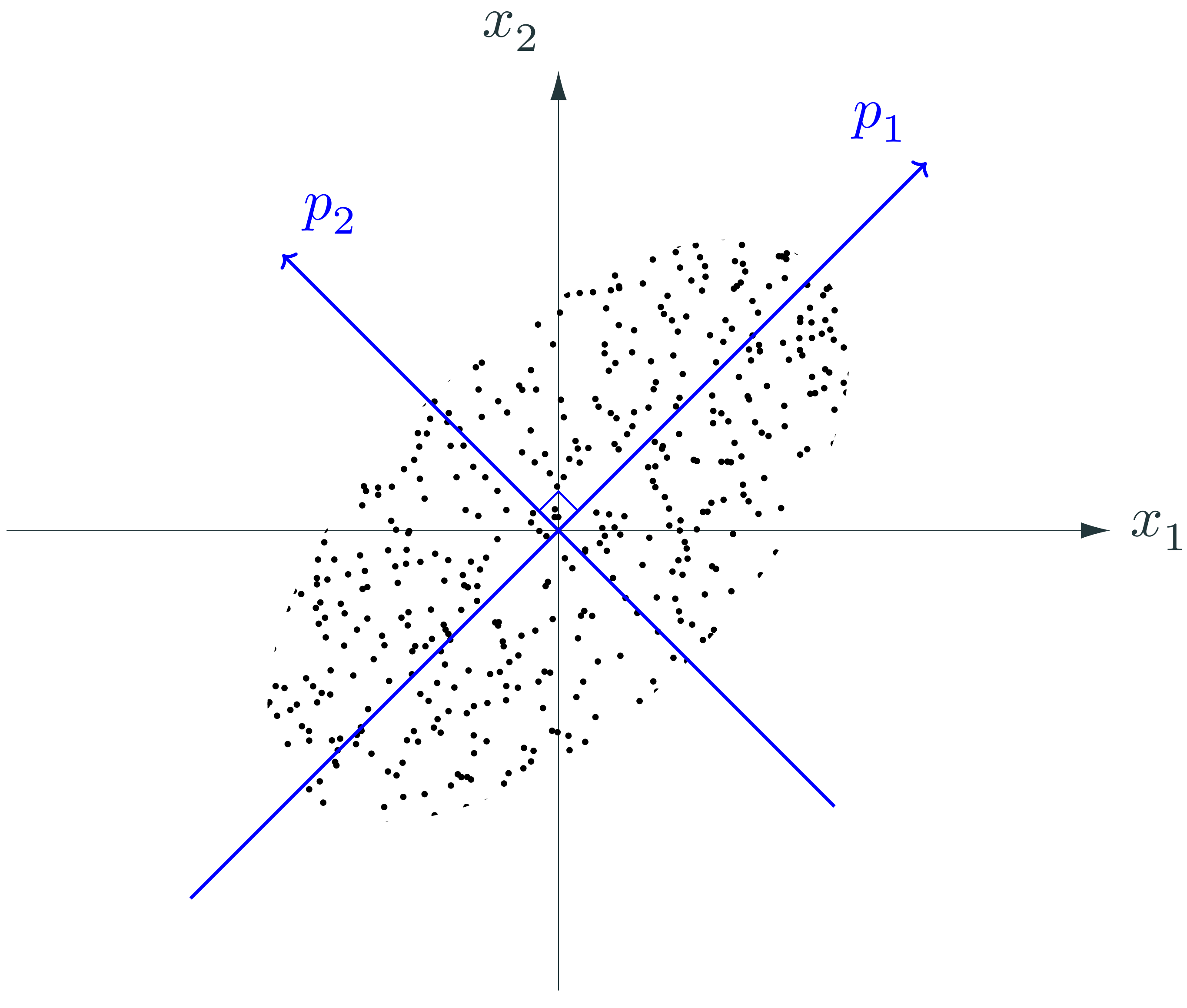
Картинка



Картинка



Картинка



И это завуалированный SVD!

Если $X = U\Sigma V^T$, то $P = XV = U\Sigma$.

И это завуалированный SVD!

Если $X = U\Sigma V^T$, то $P = XV = U\Sigma$.

Аппроксимация $\hat{X}_d$ с рангом rank $\hat{X}_d = d$ строится так:

$$\hat{X}_d = U\Sigma_d V^T,$$

где матрицу Σ_d получили из матрицы Σ оставив в ней только d штук самых больших сингулярных значений и занулив остальные.

Свойства

Главные компоненты находятся из SVD разложения, $\mathbf{p}_i = \sigma_i \mathbf{u}_i$.

Главные компоненты ортогональны.

Величины σ_i^2 равны квадратам длин $\|\mathbf{p}_i\|^2$ и пропорциональны выборочным дисперсиям $\mathbf{p}_i$.

Свойства

Главные компоненты находятся из SVD разложения, $\mathbf{p}_i = \sigma_i \mathbf{u}_i$.

Главные компоненты ортогональны.

Величины σ_i^2 равны квадратам длин $\|\mathbf{p}_i\|^2$ и пропорциональны выборочным дисперсиям $\mathbf{p}_i$.

Значение суммы

$$\|\mathbf{p}_1\|^2 + \ldots + \|\mathbf{p}_d\|^2 = \sigma_1^2 + \ldots + \sigma_d^2$$

одинаково при пошаговой и одновременной аппроксимации матрицы X.

PCA: максимизация R^2

Краткий план:

- Коэффициент детерминации.

Краткий план:

- Коэффициент детерминации.
- Максимизация R^2.

Краткий план:

- Коэффициент детерминации.
- Максимизация R^2.
- Свойства главных компонент.

Коэффициент детерминации

Рассмотрим нахождение проекции $\hat{\mathbf{y}}$ вектора $\mathbf{y}$ на линейную оболочку $\mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ столбцов матрицы X.

Коэффициент детерминации

Рассмотрим нахождение проекции $\hat{\mathbf{y}}$ вектора $\mathbf{y}$ на линейную оболочку $\mathrm{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ столбцов матрицы X.

Предположим, что среди $\mathbf{x}_i$ есть вектор-константа.

Коэффициент детерминации

Рассмотрим нахождение проекции $\hat{\mathbf{y}}$ вектора $\mathbf{y}$ на линейную оболочку Span$\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ столбцов матрицы X.

Предположим, что среди $\mathbf{x}_i$ есть вектор-константа.

Определение

Коэффициентом детерминации R^2 называют квадрат косинуса угла между центрированный вектором $\mathbf{y}'$, $y_i' = y_i - \bar{y}$, и его проекцией $\hat{\mathbf{y}}'$ на столбцы X.

$$R^2 = \cos^2 \angle(\hat{\mathbf{y}}', \mathbf{y}') = \frac{\|\hat{\mathbf{y}}'\|^2}{\|\mathbf{y}'\|^2}$$

Коэффициент детерминации

Рассмотрим нахождение проекции $\hat{\mathbf{y}}$ вектора $\mathbf{y}$ на линейную оболочку $\mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ столбцов матрицы X.

Предположим, что среди $\mathbf{x}_i$ есть вектор-константа.

Определение

Коэффициентом детерминации R^2 называют квадрат косинуса угла между центрированный вектором $\mathbf{y}'$, $y_i' = y_i - \bar{y}$, и его проекцией $\hat{\mathbf{y}}'$ на столбцы X.

$$R^2 = \cos^2 \angle(\hat{\mathbf{y}}', \mathbf{y}') = \frac{\|\hat{\mathbf{y}}'\|^2}{\|\mathbf{y}'\|^2}$$

Коэффициент детерминации — доля «объяснённого разброса».

Коэффициент детерминации

Утверждение

Коэффициент детерминации равен квадрату выборочной корреляции между исходным $\mathbf{y}$ и его проекцией $\hat{\mathbf{y}}$ на столбцы X:

$$R^2 = \rho^2(\hat{\mathbf{y}}, \mathbf{y}) = \frac{\left(\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\sum_i (y_i - \bar{y})^2 \sum_i (\hat{y}_i - \bar{y})^2}$$

Метод главных компонент

Все исходные переменные $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ предварительно стандартизируем!

Для каждого столбца $\mathbf{x}$ выполнены условия $\bar{x} = 0$, $sd(\mathbf{x}) = 1$.

На базе столбцов $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ матрицы X мы создадим $d \leq k$ новых переменных $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$.

Новые переменные будем создавать по-очереди.

Новые переменные будем называть главными компонентами.

Обозначение

Обозначим $R_i^2(\mathbf{x})$ — коэффициент детерминации при проецировании вектора $\mathbf{x}$ на линейную оболочку первых i главных компонент, Span$\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_i\}$.

Обозначение

Обозначим $R_i^2(\mathbf{x})$ — коэффициент детерминации при проецировании вектора $\mathbf{x}$ на линейную оболочку первых i главных компонент, Span$\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_i\}$.

Коэффициент $R_i^2(\mathbf{x})$ показывает, насколько хорошо первые i главных компонент предсказывают вектор $\mathbf{x}$.

Максимизация суммы R^2

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Максимизация суммы R^2

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Алгоритм

1. Компоненту $\mathbf{p}_1 = X\mathbf{v}_1$ подберём так, чтобы максимизировать сумму $R^2_1(\mathbf{x}_1) + \ldots + R^2_1(\mathbf{x}_k)$ при условии, что $\|\mathbf{v}_1\| = 1$.

Максимизация суммы R^2

Главные компоненты $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ будут линейными комбинациями столбцов X.

Алгоритм

1. Компоненту $\mathbf{p}_1 = X\mathbf{v}_1$ подберём так, чтобы максимизировать сумму $R_1^2(\mathbf{x}_1) + \ldots + R_1^2(\mathbf{x}_k)$ при условии, что $\|\mathbf{v}_1\| = 1$.
2. Компоненту $\mathbf{p}_2 = X\mathbf{v}_2$ подберём так, чтобы максимизировать сумму $R_2^2(\mathbf{x}_1) + \ldots + R_2^2(\mathbf{x}_k)$ при условии, что $\mathbf{v}_2 \perp \mathbf{v}_1$ и $\|\mathbf{v}_2\| = 1$.
3. …

И это завуалированный SVD!

Если $X = U\Sigma V^T$, то $P = XV = U\Sigma$.

Главные компоненты находятся из SVD разложения, $\mathbf{p}_i = \sigma_i \mathbf{u}_i$.

Главные компоненты ортогональны.

Величины σ_i^2 равны квадратам длин $\|\mathbf{p}_i\|^2$ и пропорциональны выборочным дисперсиям $\mathbf{p}_i$.

И это завуалированный SVD!

Если $X = U\Sigma V^T$, то $P = XV = U\Sigma$.

Главные компоненты находятся из SVD разложения, $\mathbf{p}_i = \sigma_i \mathbf{u}_i$.

Главные компоненты ортогональны.

Величины σ_i^2 равны квадратам длин $\|\mathbf{p}_i\|^2$ и пропорциональны выборочным дисперсиям $\mathbf{p}_i$.

Значение суммы

$$\|\mathbf{p}_1\|^2 + \ldots + \|\mathbf{p}_d\|^2 = \sigma_1^2 + \ldots + \sigma_d^2$$

одинаково при пошаговой и одновременной максимизации.

Смысл сингулярных значений

Квадраты сингулярных значений показывают увеличение среднего коэффициента детерминации!

$$\frac{\sigma_1^2}{\sigma_1^2 + \ldots + \sigma_k^2} = \frac{1}{k}(R_1^2(\mathbf{x}_1) + \ldots + R_1^2(\mathbf{x}_k))$$

Смысл сингулярных значений

Квадраты сингулярных значений показывают увеличение среднего коэффициента детерминации!

$$\frac{\sigma_1^2}{\sigma_1^2 + \ldots + \sigma_k^2} = \frac{1}{k}(R_1^2(\mathbf{x}_1) + \ldots + R_1^2(\mathbf{x}_k))$$

$$\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \ldots + \sigma_k^2} = \frac{1}{k}(R_2^2(\mathbf{x}_1) + \ldots + R_2^2(\mathbf{x}_k))$$

Смысл сингулярных значений

Квадраты сингулярных значений показывают увеличение среднего коэффициента детерминации!

$$\frac{\sigma_1^2}{\sigma_1^2 + \ldots + \sigma_k^2} = \frac{1}{k}(R_1^2(\mathbf{x}_1) + \ldots + R_1^2(\mathbf{x}_k))$$

$$\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \ldots + \sigma_k^2} = \frac{1}{k}(R_2^2(\mathbf{x}_1) + \ldots + R_2^2(\mathbf{x}_k))$$

$$\frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}{\sigma_1^2 + \ldots + \sigma_k^2} = \frac{1}{k}(R_3^2(\mathbf{x}_1) + \ldots + R_3^2(\mathbf{x}_k))$$

Резюме

- Жорданова нормальная форма.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.
- PCA: максимизация разброса.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.
- PCA: максимизация разброса.
- PCA: минимизация ошибки приближения.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.
- PCA: максимизация разброса.
- PCA: минимизация ошибки приближения.
- PCA: максимизация среднего R^2.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.
- PCA: максимизация разброса.
- PCA: минимизация ошибки приближения.
- PCA: максимизация среднего R^2.
- Скринкаст с SVD.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.
- PCA: максимизация разброса.
- PCA: минимизация ошибки приближения.
- PCA: максимизация среднего R^2.
- Скринкаст с SVD.
- Бонусное видео: геометрическая алгебра.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.
- PCA: максимизация разброса.
- PCA: минимизация ошибки приближения.
- PCA: максимизация среднего R^2.
- Скринкаст с SVD.
- Бонусное видео: геометрическая алгебра.

Резюме

- Жорданова нормальная форма.
- Сингулярное разложение.
- PCA: максимизация разброса.
- PCA: минимизация ошибки приближения.
- PCA: максимизация среднего R^2.
- Скринкаст с SVD.
- Бонусное видео: геометрическая алгебра.

Большое спасибо героям, прошедшим курс!

Скринкаст: SVD для снижения размерности

Это видеофрагмент с доской, слайдов здесь нет :)

Бонус: геометрическая алгебра

Это видеофрагмент с доской, слайдов здесь нет :)