Домашнее задание 4

Чэнь Сюаньдун
Группа 519/2
25 октября 2022 г.

1 SVD/Singular Value Decomposition/Сигулярное разложение

В отличие от собственного разложения, SVD не требует, чтобы матрица была квадратной.

Определение

Дана матрица A размерности m $\times$ n, тогда определяем SVD матрицы A как:

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

где U - матрица m x m, $U^TU = I$, V - матрица n x n, $V^TV = I$. Σ - матрица m x n со всеми нулями, кроме элементов на главной диагонали, каждый элемент на главной диагонали называется сингулярным значением.

Процесс SVD

. Известно, что матрица AA^T и A^TA квадратные, т.е. для них можно делать собственное разложение. Полученные собственные значения λ_i и собственные векторы удовлетворяют следующему уравнению:
Для U:$(AA^T)u_i = \lambda_i u_i$
Для V:$(A^TA)v_i = \lambda_i v_i$
Для Σ : $\mathrm{AA}^T = U\Sigma V^T V\Sigma^T U^T$ и $A^TA = V\Sigma^T U^T U\Sigma V^T$, получим что $\sigma_i = \sqrt{\lambda_i}$
Таким образом, находим U, V и Σ.

Свойство SVD

. В матрице сингулярных значений сингулярные значения упорядочены от наибольшего к наименьшему, и число сингулярных значений уменьшается особенно

быстро, во многих случаях сумма первых 10% или даже 1% сингулярных значений составляет более 99% от суммы всех сингулярных значений. Т.е., мы также можем аппроксимировать матрицу в терминах наибольших k сингулярных значений и соответствующих левого и правого сингулярных векторов. Благодаря этому важному свойству SVD может быть использована для снижения размерности (PCA).

2 PCA/Principal Component Analysis/Метод главных компонентов

PCA - это алгоритм снижения размерности.Его целью является отображение высокоразмерных данных в низкоразмерное пространство посредством некоторой линейной проекции, с расчетом на то, что данные будут наиболее информативными (с наибольшей дисперсией) в проецируемом измерении, таким образом, используя меньшее количество измерений данных при сохранении характеристик большего количества исходных точек данных. В общем, когда изучаемая проблема включает в себя несколько переменных и между ними существует сильная корреляция, мы можем рассмотреть возможность использования PCA для упрощения данных.

Процесс реализоции PCA

Предполагая, что имеется n выборок с p признаков, можно сформировать матрицу выборок X размера n*p:

$$X = \begin{bmatrix} x_1^1 & x_1^2 & x_1^3 & ... & x_1^p \\ x_2^1 & x_2^2 & x_2^3 & ... & x_2^p \\ x_3^1 & x_3^2 & x_3^3 & ... & x_3^p \\ ... & ... & ... & ... & ... \\ x_n^1 & x_n^2 & x_n^3 & ... & x_n^p \end{bmatrix} = \begin{pmatrix} X^1 & X^2 & X^3 & ... & X^p \end{pmatrix}$$

Шаг 1: Все данные минус их средние значения для лучшего последующего расчета.

Среднее значение признака: $\bar{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_j^i$, (j = 1, 2, ..., p)

Шаг 2: Найти матрицу ковариаций C

$$C = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) & ... & cov(x_1, x_p) \\ cov(x_2, x_1) & cov(x_2, x_2) & ... & cov(x_1, x_p) \\ ... & ... & ... & ... \\ cov(x_p, x_1) & cov(x_p, x_2) & ... & cov(x_p, x_p) \end{bmatrix}$$

где на главной диагонали дисперсии признаков, остальные - ковариации.

Шаг 3: Найти собственные значения и собственные векторы матрицы ковариаций C

$$Cu = \lambda u$$

Сортировать собственные значения λ в порядке убывания и выбрать k самых больших собственных значений λ , получить соответствующие им k собственных векторов u_k.

Шаг 4: Проецирование исходных признаков на выбранный вектор признаков для получения новых K-мерных признаков.

Для каждых исходного примера X^i, $-(x_1^i, x_2^i, ..., x_n^i)^T$, после проецирования это $Y^i = (y_1^i, y_2^i, ..., y_k^i)^T$:

$$Y^i = \begin{bmatrix} y_1^i \\ y_2^i \\ ... \\ y_k^i \end{bmatrix} = \begin{bmatrix} u_1^T(x_1^i, x_2^i, ..., x_n^i)^T \\ u_2^T(x_1^i, x_2^i, ..., x_n^i)^T \\ ... \\ u_k^T(x_1^i, x_2^i, ..., x_n^i)^T \end{bmatrix}$$

Определение главных компонентов

$Y^1, Y^2, ..., Y^k$ называются главными компонентами

Осуществимость

Доказательство

Чем больше ковариация, тем больше информации.

Каждый собственный вектор ковариационной матрицы представляет собой поверхность проекции, а собственное значение, соответствующее каждому собственному вектору, является дисперсией исходного признака после его проецирования на эту поверхность.

Поскольку мы хотим, чтобы как можно больше информации не было потеряно после проекции, для проецирования исходных признаков выбирается проекционная поверхность с большей дисперсией, т.е. выбираются собственные векторы с большими собственными значениями.

Исходные признаки проецируются на эти собственные векторы, спроецированные значения являются новыми собственными значениями.

Каждая поверхность проекции генерирует один новый признак, а k поверхностей проекции генерируют k новых признаков.

Геометрический смысл

Основная идея PCA заключается в отображении n-мерных признаков на k-мерные, которые являются новыми ортогональными признаками. k-мерные признаки, восстановленные из исходных n-мерных признаков, представляют собой главные компоненты.

PCA работает путем последовательного нахождения набора взаимно ортогональных осей из исходного пространства, причем выбор новых осей тесно связан с самими данными.

Первая новая ось выбирается как направление наибольшей дисперсии в исходных данных.

Вторая новая ось выбирается как направление наибольшей дисперсии в плоскости, ортогональная к первой оси.

Третья новая ось выбирается как направление наибольшей дисперсии в плоскости, ортогональная первой и второй осям

По аналогии можно получить n таких осей. С помощью новых осей, полученных таким образом, мы обнаруживаем, что большая часть дисперсии содержится в первых k осях и что последние оси содержат почти нулевую дисперсию, поэтому мы можем игнорировать остальные оси и оставить только первые k осей, которые содержат большую часть дисперсии.

Фактически, это эквивалентно уменьшению размерности данных путем сохранения только тех размерных признаков, которые содержат большую часть дисперсии, и игнорирования размерных признаков, которые содержат почти нулевую дисперсию.