Поговорим о задаче понижения размерности. Рассмотрим пару примеров. Вот есть такая выборка, у нее три размерности. При этом видно, что если просто убрать из нее признак,

который отложен по оси Z, то мы получим двумерную выборку, в которой классы будут все еще отлично разделяться между собой. Серый, желтый и красный будут разделимы даже линейными методами.

А вот более сложный случай. Здесь кажется, что оба признака — значимые. Но при этом можно заметить, что можно спроецировать эти данные на прямую, сделав их одномерными, не потеряв при этом практически никакой информации. То есть да, оба признака значимые, но при этом они линейно зависимые. И этим можно воспользоваться, чтобы устранить некоторую избыточность в данных.

Таблица для теста:

|  |  |  |  |
| --- | --- | --- | --- |
| Привет | Пока | До | Встречи |
| Увидимся | С | Тобой | В |
| Следующем | 2024 | Году | \* |

Но при этом отбора признаков здесь не хватит. Нужно сформировать новый признак на основе двух исходных.

Бывают и еще более сложные случаи, как, например, здесь. Здесь тоже видно, что можно спроецировать выборку на некоторую кривую, но при этом кривая нелинейная, и искать их довольно непросто.

Так мы приходим к задаче понижения размерности, которая состоит в формировании новых признаков на основе исходных. При этом новых признаков должно быть меньше, чем исходных, но они должны сохранять в себе как можно больше информации, которая содержалась в исходных признаках.

Одни из основных подходов к понижению размерности – это линейный подход (линейная взвешенная сумма) и метод главных компонент.

Но мы рассмотрим одним из наиболее популярных методов при решении задачи понижения размерности.

И выводили следующие обозначения: xi-тая j-тая — это значение исходного j-того признака на i-том объекте. И всего таких исходных признаков D(большое) штук.

zi-тая j-тая — это значение нового j-того признака на i-том объекте. И всего таких новых преобразованных признаков d (маленькое) штук.

В линейном подходе zi-тая j-тая, то есть значение j-того признака на i-том объекте линейно выражается через исходные признаки. То есть мы суммируем исходные признаки xi-тая j-тая с весами wj-тая k-тая. На самом деле эту формулу можно переписать в матричном виде.

Для этого немного поработаем над ней. Поменяем местами множители wj-тая k-тая и xi-тая j-тая, что всегда можно сделать без потерь в этой формуле, и попробуем привести это к формуле умножения матриц. Для этого перейдем от матрицы wj-тая k-тая к транспонированной. Тогда индексы будут k и j, w транспонированная k-тая j-тая. А это уже формула перемножения матриц x и w.

То есть получается, что матрица новых признаков Z вычисляется как произведение исходной матрицы объекты-признаки X и транспонированной матрицы W.

Потребуем, чтобы матрица весов W была ортогональной. То есть чтобы произведение W транспонированного на W равнялась единичной матрице.

На самом деле это ограничение никак не будет нам мешать. Если б мы начали решать задачу без него, то оказалось бы, что нам нужно ввести некоторое подобное ограничение, чтобы решение было единственным.

Итак, если данное требование будет выполнено, то из формулы Z = X \* W транспонированное можно получить формулу для X, поскольку матрица W ортогональна. X будет вычисляться как Z \* W.

Понятно, что это равенство вряд ли можно выполнить точно. Поскольку в матрице Z будет меньше признаков, чем в X, то если ранг матрицы X больше, чем d, больше, чем число новых признаков, то данное равенство точно нельзя будет выполнить строго.

В этом случае будем требовать, чтобы отклонение исходной матрицы признаков X от восстановленной матрицы X с помощью произведения Z на W было как можно меньше, чтобы эти матрицы были как можно сильнее похожи друг на друга. Размер этого отклонения будем вычислять с помощью нормы Фробениуса разности матриц X и Z \* W. Норма Фробениуса матрицы — это просто сумма квадратов ее значений, аналог l2 норма для вектора. Будем минимизировать данную норму. По сути, это задача матричного разложения. Мы хотим представить матрицу X в виде произведения двух матриц Z и W, которые будут иметь меньший ранг. То есть мы хотим уменьшить ранг матрицы, при этом потеряв как можно меньше информации в ней.

Есть и немного другой подход к постановке задачи метода главных компонент.

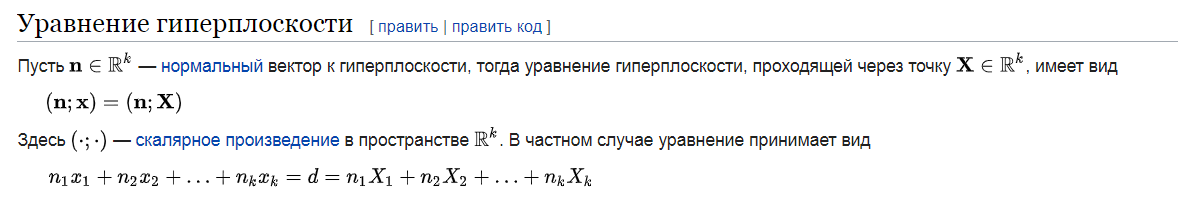
Представьте, что у нас есть вот такая выборка, и мы хотим спроецировать ее на некоторую прямую. В этом случае прямая будет тем лучше, чем меньше будет ошибка проецирования. Под ошибкой проецирования мы понимаем сумму по всей выборке расстояний от объекта до его проекции на эту прямую. Понятно, что чем меньше эти расстояния, тем лучше прямая приближает данные, тем меньше будет ошибка и тем больше информации мы сохраним. В идеальном случае мы хотели бы провести прямую так, чтобы она уже проходила через все объекты выборки. Но понятно, что в данном случае это невозможно. Поэтому будем искать такую прямую, расстояние от которой до всех объектов в сумме будет минимальным. В общем случае, когда признаков много, мы будем пытаться проецировать выборку не на плоскость, не на прямую, а на гиперплоскость.

Можно показать, это известно из аналитической геометрии, что есть два способа задания гиперплоскости.

Первый — с помощью вектора нормали. Мы его использовали, например, в линейных методах.

Второй — с помощью направляющих векторов. Если размерность пространства D, и мы строим в нем гиперплоскость, то если выбрать на этой плоскости D линейно независимых векторов, которые лежат в ней, то они тоже будут однозначно задавать эту плоскость, эту гиперплоскость. И если эти векторы, направляющие векторы, составить в матрицу W, так что каждый столбец этой матрицы, это один направляющий вектор, то опять же можно показать, что проекция точки xi-тая на данную гиперплоскость будет вычисляться по формуле xi-тая \* W. Тогда мы будем, чтобы минимизировать ошибку проецирования на гиперплоскость, мы будем минимизировать сумму норм отклонений xi-тых исходных объектов, от их проекции xi-тая на W.

На самом деле у этой формулы есть некоторые проблемы. xi-тая на w — это вектор немного другой размерности, который задает... задается координатами в проекционных векторах. Эту проблему можно решить, если перейти немножко к другому базису, но не будем об этом. Нам сейчас важна идея. Можно решать задачу путем минимизации отклонений объектов от проекционной гиперплоскости.

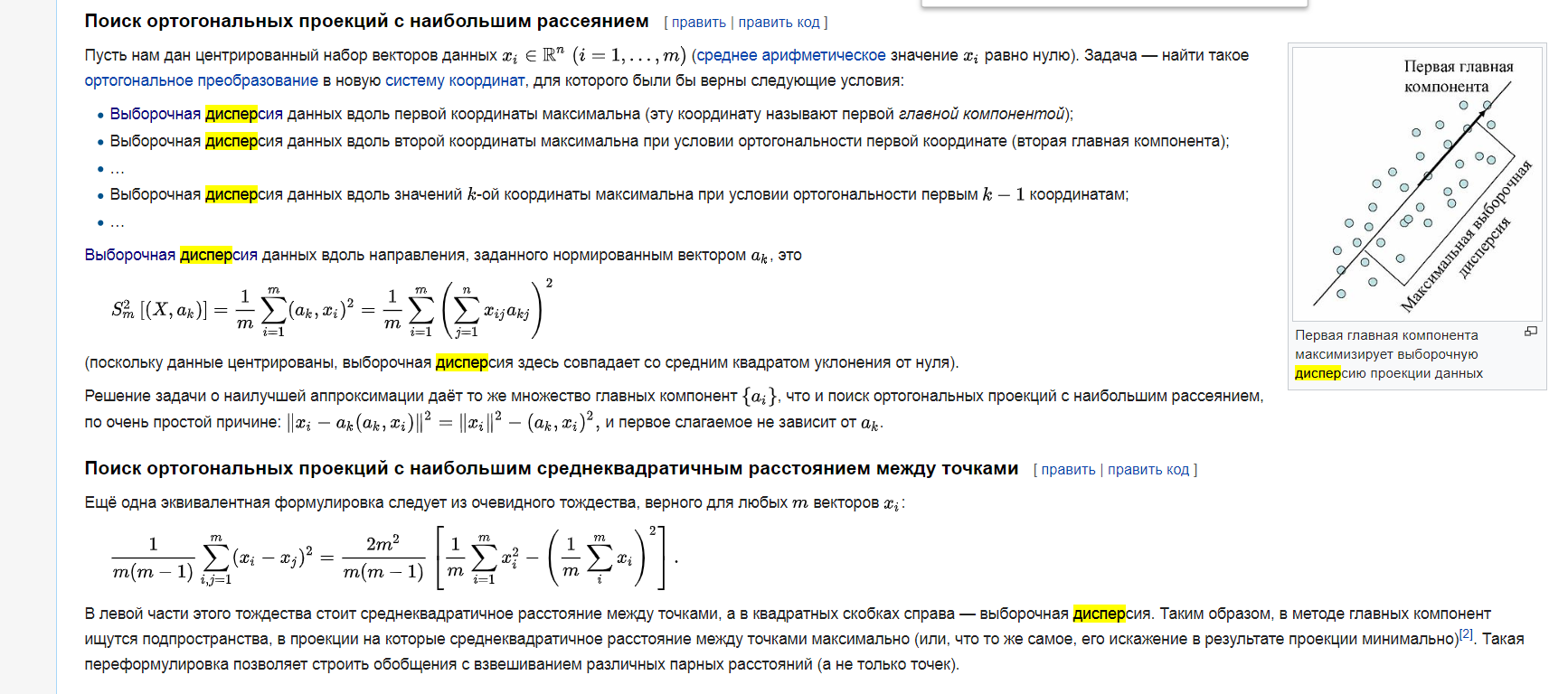


Есть и третий взгляд на метод главных компонент. Представьте, что у нас есть вот такая выборка.И мы хотим снова выбрать прямую, на которую будет оптимально ее спроецировать. Понятно, что зеленая прямая будет лучше, поскольку при проецировании на нее сохранится гораздо больше информации выборки.

Как формализовать понятие информации? Например, можно говорить о дисперсии. Чем больше будет дисперсия выборки после проецирования на прямую, тем лучше, тем больше информации мы сохранили.

Собственно, для данного случая этот критерий хорошо подходит.

Дисперсия выборки после проецирования на зеленую прямую будет гораздо больше, чем после проецирования на феолетовую прямую. Формально дисперсию выборки после проецирования можно записать с помощью вот такой формулы, которая представляет собой сумму по всем новым признакам, произведение столбцов матриц весов wj-тое и матриц объекты-признаки для исходного набора признаков. Чем больше значение этой суммы, тем больше дисперсия выборки после проецирования на гиперплоскость, которая задается матрицей весов w. Таким образом, данное выражение нужно максимизировать. Чем оно будет больше, тем больше информации выборки мы сохраним после проецирования.



Давайте разберемся, как решать задачу метода главных компонент.

одна из постановок метода главных компонент — это максимизация дисперсии. Есть некоторое облако точек, где каждая точка — это объект выборки, и нужно найти такие оси, при проецировании на которые мы сохраним как можно больше дисперсии исходной выборки. Чем больше дисперсии мы сохраним, тем больше информации останется после понижения размерности.

Формально задача максимизации дисперсии после проецирования записывается вот так.

В первой строке записана собственно дисперсия после проецирования — она выражается через направление wi, а поскольку мы хотим оставить как можно больше информации, дисперсию нужно максимизировать. Также мы вводим органичение, что матрица весов W должна быть ортогональной — это нужно, чтобы решение было единственным.

В методе главных компонент есть один нюанс: выражение, через которое мы записали дисперсию, будет означать именно дисперсию выборки только в том случае, если матрица объекты-признаки центрирована, то есть если среднее каждого признака равно нулю.

Поэтому будем считать, что, собственно, выборка центрирована, что мы уже вычли среднее из каждого столбца в матрице объекты-признаки.

Итак, чтобы разобраться, как устроено решение этой задачи, давайте попробуем сначала разобраться с простым частным случаем — случаем, когда мы хотим найти ровно одну компоненту, на которую будем проецировать всю выборку, и хотим выбрать ее так, чтобы дисперсия после проецирования была максимальной.

Задача будет выглядеть вот так.

Вектор весов, который характеризует направление — это w1, и он умножается на матрицу X транспонированное, это умножается на матрицу X и снова на вектор весов w1. Это мы хотим максимизировать.

Ограничение же вырождается в то, что L2 норма этого вектора весов должна равняться единице.

Итак, как вы знаете, чтобы решать такие задачи условной оптимизации, нужно выписать их лагранжиан. Лагранжиан будет выглядеть вот так.

Это собственно критерий, который мы максимизируем минус λ умножить на ограничение.

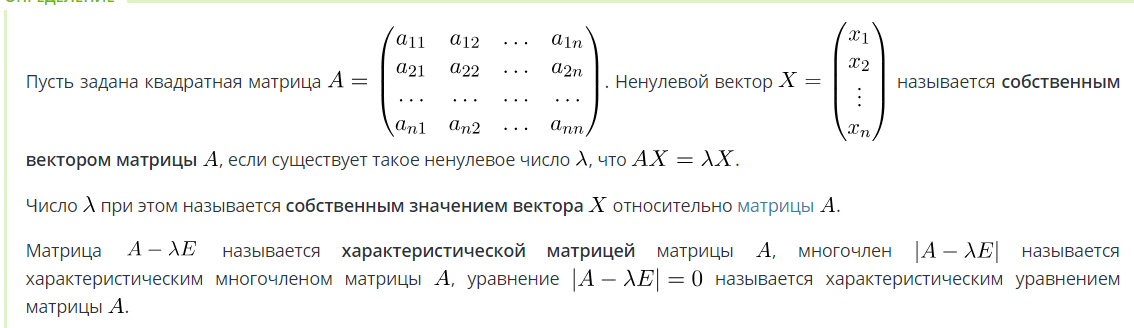
λ умножить на w1 транспонированное на w1 минус единица.

Далее этот лагранжиан нужно продифференцировать по тому, что мы хотим найти, то есть по w1. Если воспользоваться формулами матричного дифференцирования, можно получить вот такую формулу.

C:\Users\Dimasik\Desktop\Интеллектуальные системы\Лекции\Лекция 2\Снимок.PNG

Если там ее немного преобразовать, перенести одно слагаемое в правую часть и поделить на 2, то мы получим вот такое уравнение. X транспонированное на X умножить на w1 равняется λ на w1.

Давайте обратим внимание, что что говорит это уравнение? Оно говорит, что вектор w1 является собственным вектором матрицы X транспонированное на X, поскольку при умножении матрицы на этот вектор, мы получаем этот же вектор, умноженный на некоторое число λ, и число λ является собственным значением, соответствующим этому собственному вектору.



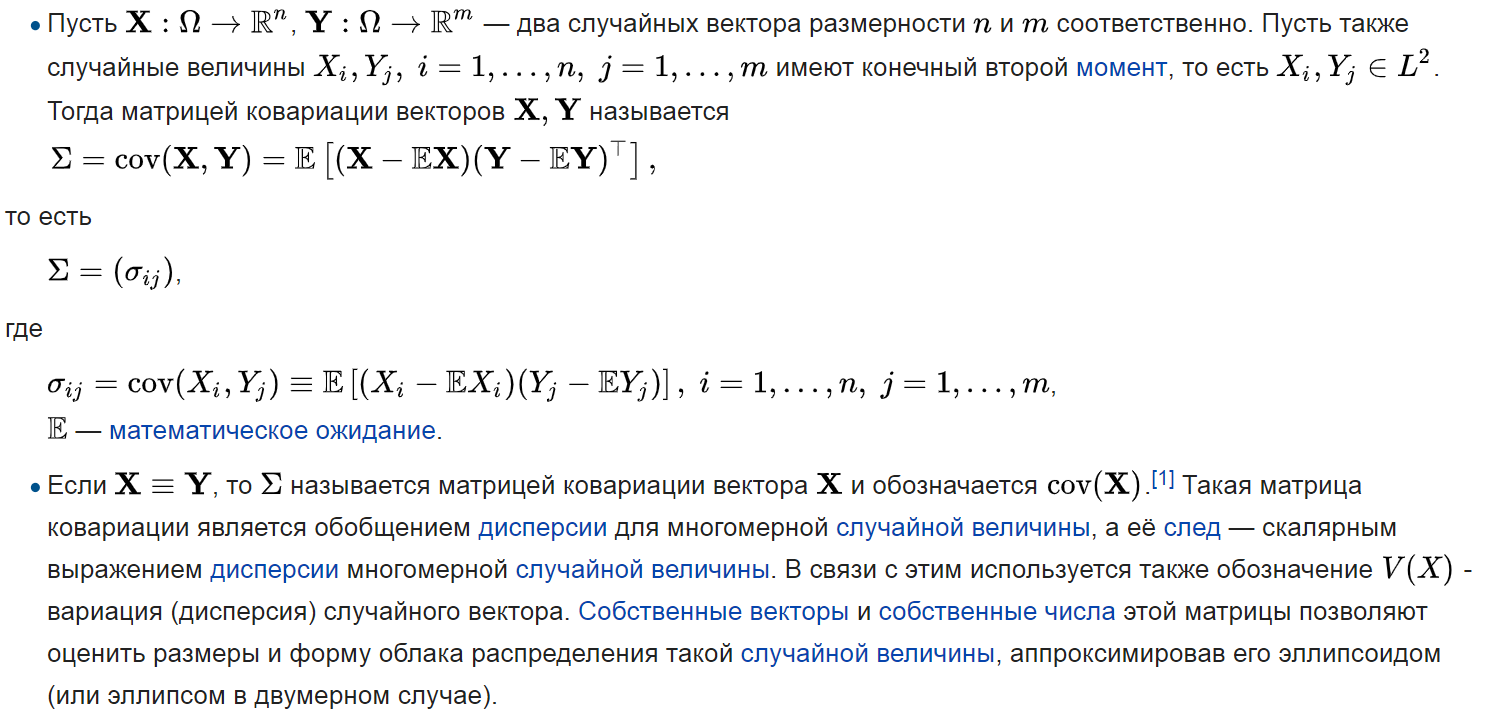
Если мы подставим полученное нами условие в функционал задачи, то обнаружим, что дисперсия выборки после проецирования будет равна как раз λ, то есть будет равна собственному значению, соответствующему собственному вектору, который мы выбрали.

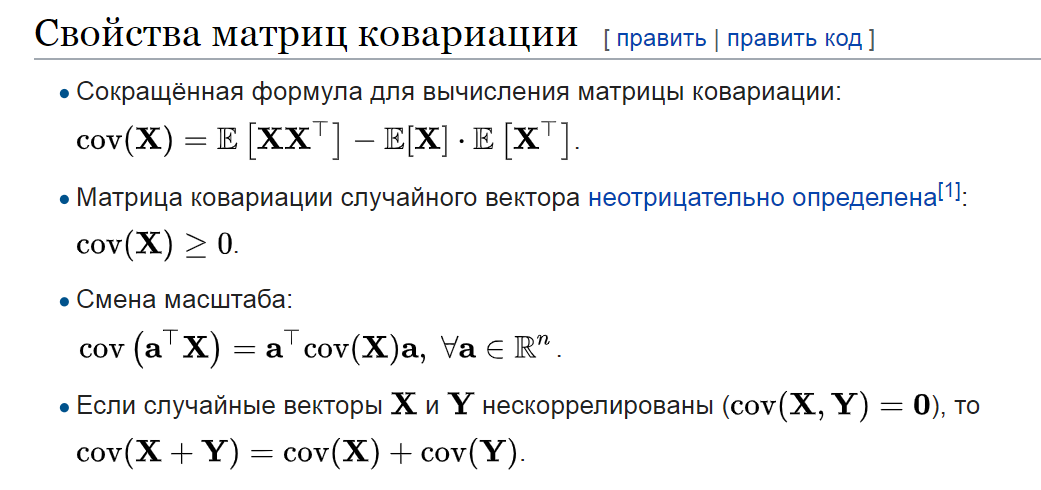
Таким образом, поскольку мы хотим максимизировать дисперсию, нам нужно выбирать максимальное собственное значение и собственный вектор,который соответствует этому максимальному собственному значению.

Итак, мы получаем, что первая компонента в методе главных компонент — это собственный вектор матрицы X транспонированное на X, который соответствует максимальному собственному значению этой матрицы.

При чем, обратите внимание, что X транспонированное на X — это матрица ковариации, то есть именно та матрица, которая характеризует дисперсию нашей выборки. Неудивительно, что именно через ее характеристики выражается решение метода главных компонент.







Визуально это выглядит вот так: есть облако точек, и мы выбираем именно то направление, при проецировании на которое мы сохраним как можно больше дисперсии. И, собственно, это направление будет как раз равно первому собственному вектору матрицы ковариации.

Если мы продолжим наши выкладки и будем искать оптимальное второе направление, третье и так далее, то обнаружим, что, например, w2 — второе направление в методе главных компонент — соответствует собственному вектору матрицы X транспонированное на X, соответствующему второму по величине собственному значению, и так далее. Кстати, отсюда же можно получить, что через собственные значения можно выразить долю дисперсии, которую мы сохранили.

Дисперсия всей исходной выборки — это сумма всех собственных значений матрицы X транспонированное на X, то есть сумма от 1 до D; дисперсия выборки после проецирования на d главных компонент — это сумма максимальных d собственных значений.

Таким образом, с помощью вот такой дроби можно понять, какую долю дисперсии мы сохранили, спроецировав нашу выборку на главные компоненты.

При решении задачи метода главных компонент очень пригождается сингулярное разложение.

Сингулярное разложение матрицы X представляет эту матрицу в виде произведения трех других матриц: U умножить на D умножить на V транспонированное. При этом U и V — это ортогональные матрицы, а D — диагональная матрица. Столбцы матрицы U — это собственные векторы матрицы X на X транспонированное, столбцы матрицы V — это собственные векторы матрицы X транспонированное на X, а на диагонали матрицы D стоят собственные значения этих обоих матриц, которые, оказывается, совпадают.

Кстати, эти собственные значения матриц X транспонированное на X и X на X транспонированное называются сингулярными числами.

Итак, мы приходим к следующему алгоритму для решения задачи метода главных компонент.

Мы находим сингулярное разложение матрицы X, формируем матрицу весов W из собственных векторов, из столбцов матрицы V, соответствующих максимальным сингулярным числам, и после этого можем делать преобразования.

Чтобы спроецировать объекты, записанные в матрицу X на наши главные компоненты, мы просто умножаем матрицу X на W и получаем матрицу Z, которая является матрицей объекты-признаки для нового сокращенного признакового описания.

Итак, мы с вами разобрались, как решать задачу метода главных компонент. Чтобы ее решить, нужно найти собственные векторы матрицы ковариации X транспонированное на X и выбрать те из них, которые соответствуют максимальным собственным значениям. Именно проецирование на эти векторы будет позволять сохранить как можно больше дисперсию. Также мы выяснили, что через собственные значения этой матрицы можно выяснить, какую именно долю дисперсии мы сохранили при таком понижении размерности.