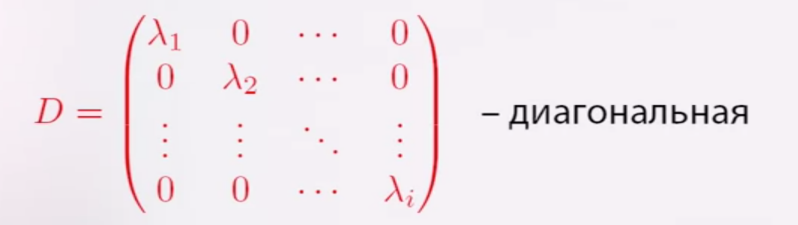
**Разложение матриц в произведение, сингулярное разложение**

Идея матричных разложений очень проста. Иногда бывает удобно представить матрицу как произведение каких-то других матриц, обладающих некоторыми интересными свойствами.

*X = ST\*D\*S  
S – ортогональная*



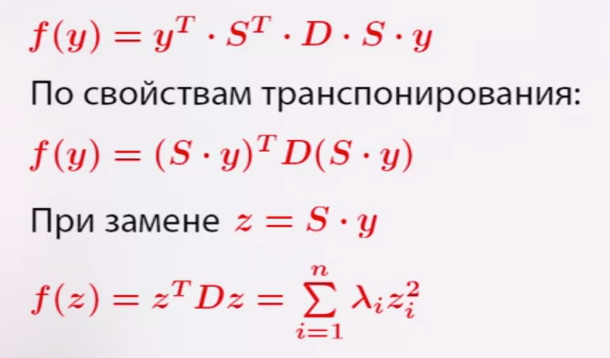
Первый пример — спектральное разложение матрицы.

Если матрица X симметрична, то ее можно представить в виде произведения S (транспонированное) \* D \* S, где матрица D диагональная, а матрица S ортогональная.  Все матрицы имеют тот же размер, что и матрица X. Элементы диагональной матрицы неотрицательные и являются собственными числами матрицы X.

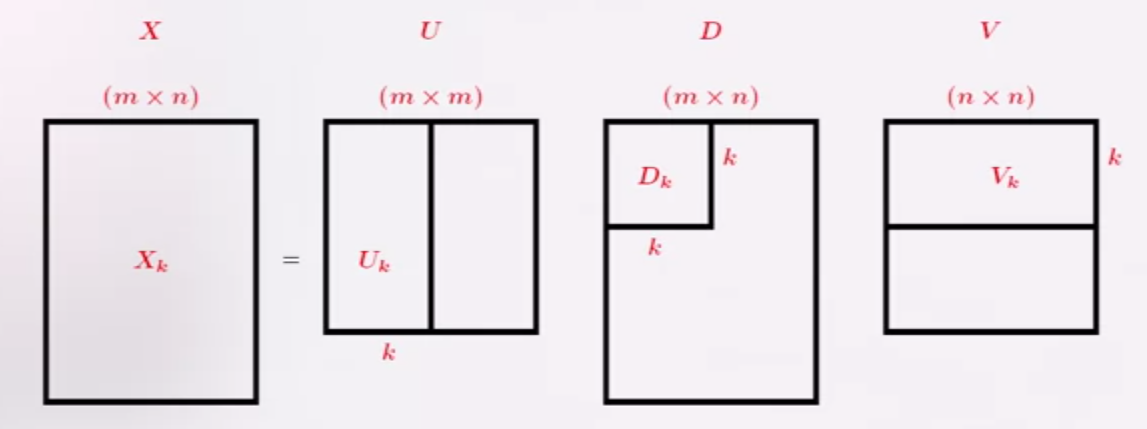
*f(y) = yT\*X\*y*  
  
Ну, например, часто встречаются функции вида f(y) = y (транспонированное) \* X \* y, где y — некоторый вектор.

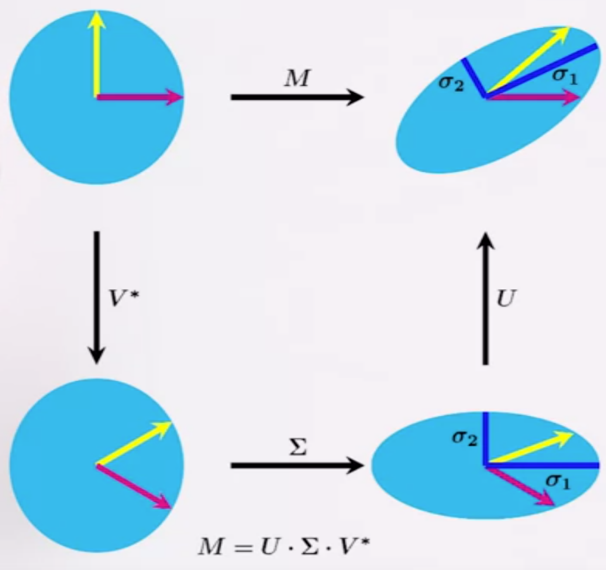
Такие выражения называются **квадратичными формами**,

и вы можете их встретить, например, при анализе многомерного нормального распределения или при анализе матрицы ковариации.



Оказывается, что если мы воспользуемся спектральным разложением матрицы, то мы сможем провести некоторую очень простую замену переменной, введя вместо переменной y переменную z = S \* y, и теперь квадратичная форма будет иметь намного более простой вид. Это будет просто взвешенная сумма квадратов координат новой переменной.



Следующий пример — это сингулярное разложение матрицы. В этом случае мы представляем матрицу X как произведение трех матриц: U \* D \* V. Матрицы U и V — ортогональные, матрица D — диагональная. Но теперь мы не требуем от матрицы X симметричности. Геометрический смысл здесь довольно прост. Исходная матрица задает некоторое линейное преобразование, и матрицы, которые используются в произведении, тоже задают некоторые преобразования. И мы просто вместо того чтобы рассматривать какое-то одно сложное преобразование, представляем его как сначала некоторое ортогональное преобразование, то есть какой-нибудь поворот, инверсии, а затем растяжение вдоль осей с помощью диагональной матрицы и затем снова поворот.

Применений у сингулярного разложения масса.

На самом деле очень часто, выводя какие-то выражения для разных методов машинного обучения, мы будем сталкиваться с тем, что подставить сингулярное разложение — это очень хорошая идея, которая приводит к каким-нибудь способам избежать разные сложные ситуации.

**Приближение матрицей меньшего ранга**

Если более формально, то ранг матрицы — это максимальное

количество линейно независимых столбцов или строк матрицы. Можно ввести ранг и по-другому.

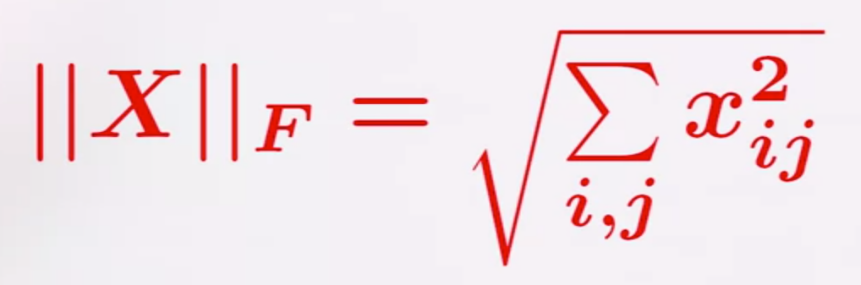
Можно сказать, что ранг — это максимальный размер подматрицы с ненулевым определителем. Также полезно заметить, что если размер матрицы X — m на n, то ранг матрицы будет точно не больше, чем минимум из этих двух чисел, что в целом понятно из определения.

*X = A\*B*

*X - m×n, A - m×k, B - k×n*

В случае, если мы рассматриваем произведение двух матриц,

ну например, матрица A имеет размер m на k, матрица B имеет размер k на n, то их произведение конечно имеет размер m на n. Однако, если число k очень маленькое, ну то есть меньше m и n, то ранг этого произведения тоже не может быть больше, чем k.

Итак, зачем нам может быть нужно приближать матрицу матрицей меньшего ранга? А у нас может быть предположение, что матрица, которую мы видим в действительности, не совсем правильная.   
  
*X* ≈ *X’ = UVT*  
  
Ну например, это могут быть какие-нибудь рейтинги фильмов или какие-нибудь признаки. И мы можем думать, что там есть какой-то шум. И нам хочется восстановить более простую матрицу, которая будет достаточно близка к той, которую мы пронаблюдали. Один из способов это делать заключается в том, что мы не просто пытаемся подобрать матрицу меньшего ранга k, а мы представляем ее как произведение двух матриц U и V транспонированное таких размеров, чтобы ранг произведения был не больше, чем k.   
  
*X* ≈ *X’ = UVTU - m×k, V - n×k  
||X - UVT||→ min*  
  
Но что значит «приблизить»?   
Что значит, что X будет примерно равно U умножить на V транспонированное? Ну просто нам нужно ввести некоторую норму на матрицах, сказать, что норма разности X − U \* V транспонированное должна быть минимальной и подбирать матрицы U и V таким образом.   
  
  
  
Как ввести норму матрицы?   
На самом деле есть много способов. Один из них — это норма Фробениуса. Выглядит достаточно интуитивно. Нужно просто взять и просуммировать квадраты всех элементов матрицы и извлечь из этого всего корень.

*Пример 1: преобразование признаков*.

Если исходная матрица X — это матрица с признаками некоторых объектов, то новая матрица U содержит столько же строчек, сколько и в ней, и тоже может интерпретироваться как матрица признаков объектов. При этом, если число k у нас намного меньше, чем n (исходное количество признаков),

то мы еще и выполняем задачу понижения размерности пространства признаков. Так как по матрице U можно довольно неплохо восстановить матрицу X с помощью

матрицы V, то мы выполняем задачу понижения размерности таким образом, что теряем как можно меньше полезной информации. Ну то есть практически все нужное нам сохраняем.

*Другой пример — задача рекомендаций.*

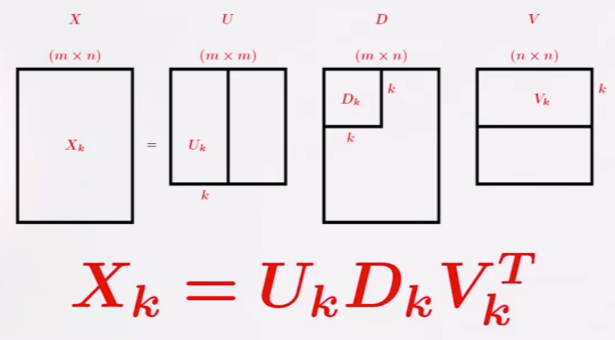
Пусть X — это матрица с оценками xᵢⱼ, поставленными пользователем i фильму j. Какие-то значения матрицы нам известны, а какие-то остаются для нас тайной.   
  
Возьмем простую модель.   
Будем приближать xᵢⱼ произведением uᵢ на vⱼ, где uᵢ отражает интересы пользователя, а vⱼ отражает некоторые признаки фильма. Возникает следующая идея: настроить u и v таким образом, чтобы хорошо прогнозировать известные значения xᵢⱼ, а неизвестные значения спрогнозируем как произведение уже настроенных векторов признаков.   
Тогда задача немного модифицируется. Теперь мы берем сумму не по всем элементам ij, а только по тем, которые нам известны.  
  
**Связь сингулярного разложения и приближения матрицей меньшего ранга**

Сингулярное разложение матриц — это представление исходной матрицы произведением трех других: ортогональной, диагональной и снова ортогональной.

Оказывается, что если мы рассматриваем матрицу Xk,

которая представляет из себе произведение трех матриц,

получаемых из SVD, она будет обладать некоторыми особыми свойствами.



Что же это за три матрицы?

Из первой матрицы, из SVD, возьмем k столбцов первых, из второй матрицы возьмем квадрат размера k × k, он должен соответствовать самым большим диагональным числам, и из третьей матрицы возьмем k строк.

И вот эти вот матрицы и перемножим, получив матрицу Xk.

Так что же за удивительное свойство, которым она будет обладать? Оказывается, матрица Xk будет наилучшим приближением исходной матрицы X по норме Фробениуса. Это совершенно нетривиальный на первый взгляд факт, но его можно с успехом использовать в разных прикладных ситуациях.

Таким образом, SVD позволяет нам сделать наилучшее приближение исходной матрицы по норме Фробениуса. В рекомендательных системах, используя матричные разложения, мы тоже ищем наилучшее приближение.

Наверняка это неспроста. Есть такой вариант: давайте возьмем SVD, как-то распределим диагональную матрицу между двумя другими и будем получившийся результат использовать как матричное разложение, например в рекомендательных системах.

Первый недостаток такой мысли в том, что SVD не то чтобы очень просто сделать.

Второй недостаток заключается в том, что SVD мы делаем для всей матрицы, а некоторые значения нам неизвестны, значит вместо них нам все равно придется подставить какие-то искусственные значения — нули или среднее значение.

Однако так сложилось, что очень долго в рекомендациях любое матричное разложение (на две матрицы, не на три) называли SVD. Ну, видимо, из-за вот этого свойства SVD.