Конспектировала: Ермолова Марина.

1. Задачи из домашнего задания

1.1. Задача 1.13

Вспомним определения:

$$TSS = \sum_i (y_i - \overline{y})^2$$

$$RSS = \sum_i (y_i - \hat{y}_i)^2$$

$$ESS = \sum_i (\hat{y}_i - \overline{y})^2$$

1. Преположим, что после добавления точки к регрессии RSS уменьшится. Тогда получается, что

$$\sum_{i=1}^{n} (y_i - \hat{y}_i^*)^2 \le \sum_{i=1}^{n+1} (y_i - \hat{y}_i^*)^2 < \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Из этого следует, что новая регрессия обеспечивает меньшую сумму квадратов, а значит исходная регрессия не была оптимальной. Противоречие.

$$RSS \le RSS^*$$

2. Проведем рассуждения, аналогичные прошлому пункту: Допустим, $TSS \le TSS^*$

$$\sum_{i=1}^{n} (y_i - \overline{y}^*)^2 \le \sum_{i=1}^{n+1} (y_i - \overline{y}_i^*)^2 < \sum_{i=1}^{n} (y_i - \overline{y}_i)^2$$

Известно, что если проецировать любой вектор на $\vec{1}$, то получается $\overline{y}\vec{1}$. Но мы знаем, что при решении задачи минимизации

$$\min_a \sum (y_i - a)^2$$

$$a_{opt} = \overline{y}$$

А по нашему предположению $\overline{y}^*$ дает минимум меньше, чем мы получаем, оптимизируя. Возникает противоречие.

3. ESS может меняться как угодно. Попробуем решить по аналогии с предыдущими пунктами:

$$\sum_{i=1}^{n} (\hat{y}_i{}^* - \overline{y}^*)^2 \le \sum_{i=1}^{n+1} (\hat{y}_i{}^* - \overline{y}_i^*)^2 < \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$$

Видим, что провести аналогичные рассуждения мы не можем.

а) ESS может падать. Например, Например, можно разместить точку настолько низко, чтобы прямая стала горизонтальной. Тогда все $\hat{y}_i = \overline{y}^*$ и $ESS = 0$.

б) ESS может возрастать. Расположим $\hat{y}_i$ на прямую, тогда $\overline{y}$ не меняются, $\overline{y}$ вырос, появилось новое слагаемое, из-за которого ESS возрастает.

2. Сингулярное разложение матрицы SVD

Сингулярное разложение матрицы используется при снижении размерности и не только.

2.1. Смысл условия $U^T U = I$

Вспомним смысл условия $U^T U = I$, где I – единичная матрица.

1. $$U^T U = I \Leftrightarrow U^{-1} = U^T \Leftrightarrow U U^T = I$$

2. $u_1, \ldots, u_n$ – вектор-столбцы матрицы U.

$$\begin{bmatrix} \ldots & u_1^T & \ldots \\ \ldots & u_2^T & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & u_n^T & \ldots \end{bmatrix} \begin{bmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & & u_n \\ \vdots & \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

$$u_2^T \cdot u_2 = 1$$

$$u_2^T \cdot u_7 = 0$$

Следовательно, столбцы u имеют единичную длину и ортогональны друг другу.

2.2. Свойства ортогональной матрицы

$$U \cdot x = \tilde{x}$$

1. Матрица U сохраняет длину вектора.
   $\tilde{x}^T \tilde{x} = (Ux)^T (Ux) = x^T U^T U x = x^T I x = x^T x$

2. Матрица U сохраняет угол.
   $U \cdot x = \tilde{x}$
   $U \cdot y = \tilde{y}$
   $\cos(\tilde{x}, \tilde{y}) = \dfrac{\tilde{x}^T \tilde{y}}{\|\tilde{x}\| \|\tilde{y}\|} = \dfrac{x^T U^T U y}{\|x\| \|y\|} = \dfrac{x^T y}{\|x\| \|y\|} = \cos(x, y)$

U может быть, например, поворотом или отражением.

2.3. Идея SVD геометрически

X - матрица переменных $n \times k$, k – количество переменных, n – количество наблюдений. $X \cdot a = b$
X превращает k - мерные вектора в n-мерные.
Любое действие X при выборе удачного базиса в $\mathbb{R}^k$ и при выборе удачного базиса в $\mathbb{R}^n$ очень просто действует.

Допустим, в пространестве R^k мы выбрали базис $v_1, v_2, \ldots, v_k$. А в R^n мы выбрали базис $u_1, u_2, \ldots, v_k$. Оказывается, можно так подобрать базисы, что:

$$\begin{cases} Xv_1 = \lambda_1 u_1 \\ Xv_2 = \lambda_2 u_2 \\ \ldots \end{cases}$$

Так как в общем случае $n \neq k$, то, либо в некоторые u не переходят v, либо если $k > n$, $Xv_j = 0, j > n$.

Пример. Проецирование $X : \mathbb{R}^3 \to \mathbb{R}^2$
С помощью сингулярного разложения независимо от действия можно выбрать хороший базис в $\mathbb{R}^2$ и $\mathbb{R}^3$, что действие будет очень просто. Выберем базис:

$$\begin{cases} Xv_1 = u_1 \\ Xv_2 = u_2 \\ Xv_3 = 0 \end{cases}$$

Пример: Поворот на плоскости по часовой стрелке на 45 градусов $X : \mathbb{R}^2 \to \mathbb{R}^2$
Выберем базис:

$$\begin{cases} Xv_1 = u_1 \\ Xv_2 = u_2 \end{cases}$$

2.4. Представление матрицы в виде произведения

Любую матрицу X $n \times k$ можно представить в виде $X = U \cdot \Sigma \cdot V^T$, где:
$U^T \cdot U = I_{n \times n}$
$V^T \cdot V = I_{k \times k}$, т.е. V не меняет углы и расстояния. а просто меняет базис.
Σ - диагональная матрица, необязательно квадратная. Например,

$$\Sigma = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Пусть x_V - столбец координат в базисе V, а x - столбец координат в исходном базисе, тогда

$$Vx_V = x$$

$$x_V = V^T x$$

2.5. Задача 5.14

$X = U \begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} V^T$. Найти сингулярное разложение матриц:

1. X^T

2. $10X$

3. X^TX

4. XX^T

Решение:

1. $X^T = \left(U\begin{pmatrix}7 & 0 & 0\\ 0 & 3 & 0\end{pmatrix}V^T\right)^T = V\begin{pmatrix}7 & 0 & 0\\ 0 & 3 & 0\end{pmatrix}^T U^T = V\begin{pmatrix}7 & 0\\ 0 & 3\\ 0 & 0\end{pmatrix}U^T$

2. U, V – матрицы, сохраняющие длину, поэтому $\Sigma = \begin{pmatrix}70 & 0 & 0\\ 0 & 30 & 0\end{pmatrix}$

   В итоге:
$$10X = U\begin{pmatrix}70 & 0 & 0\\ 0 & 30 & 0\end{pmatrix}V^T$$

3.
$$X^TX = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T$$
$$X^TX = V\begin{pmatrix}7 & 0\\ 0 & 3\\ 0 & 0\end{pmatrix}\begin{pmatrix}70 & 0 & 0\\ 0 & 30 & 0\end{pmatrix}V^T = V\begin{pmatrix}49 & 0\\ 0 & 9\end{pmatrix}V^T$$

4.
$$XX^T = U\Sigma V^TV\Sigma^TU^T = U\Sigma\Sigma^TU^T$$

$$X^TX = U\begin{pmatrix}70 & 0 & 0\\ 0 & 30 & 0\end{pmatrix}\begin{pmatrix}7 & 0\\ 0 & 3\\ 0 & 0\end{pmatrix}U^T = U\begin{pmatrix}49 & 0 & 0\\ 0 & 9 & 0\\ 0 & 0 & 0\end{pmatrix}U^T$$

2.6. Алгоритм SVD разложения компьютером

1. $X = UBV^T$, где U,V – отражение, а B - бидиогональная матрица.
$$B = \begin{pmatrix}a_{1,1} & a_{1,2} & 0 & \dots & 0\\ 0 & a_{1,1} & a_{1,2} & \dots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \dots & a_{n-1,n-1} & 0\\ 0 & \dots & 0 & a_{n,n-1} & a_{n,n}\end{pmatrix}$$

   На первом этапе побираются такие отражения, чтобы получить бидиагональную матрицу. Этот этап условно точный, можем его выписать численно.

2. Этот этап численный. С помощью вращений $X = U\Sigma V^T$ достигаем конечного результата с какой-то точностью.

Важно, что на компьютере при сингулярном разложении X никогда не считается X^T.

2.7. SVD разложение руками

$$X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} = U\Sigma V^T$$

$$X^T X = V\Sigma^T \Sigma V^T == \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$XX^T = U\Sigma\Sigma^T U^T = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

$$\det \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = 0 \to \lambda = 3, 1$$

Собственные вектора: $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Нормированная матрица: $V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$

Тогда

$$X^T X = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}^T$$

Собственные числа для XX^T

$$\det \begin{pmatrix} 2-\lambda & 1 & -1 \\ 1 & 1-\lambda & 0 \\ -1 & 0 & 1-\lambda \end{pmatrix} = 0 \to \lambda = 0, 1, 3$$

Собственные вектора: $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$

Нормированная матрица U

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

В итоге получаем разложение:

$$X = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

3. Метод главных компонент PCA

3.1. Задача

Наша задача снизить размерность матрицы так, чтобы несильно двигать точки. Идея: есть две координаты. Возьмем двумерное облако точек. Хотим несильно двигая точки сделать так, чтобы точки лежали на одной прямой.

Пусть x – i-ая строка-столбец в матрице X или столбец в матрице X^T. Тогда наша задача сводится к

$$\min_{\mu,V,\lambda_i} \sum ||x_i - \mu - V\lambda_i||^2$$

x_i размера $k \times 1$, μ размера $k \times 1$, V размера $k \times p$, λ_i размера $p \times 1$. В столбцах V находятся $k-$мерные ортогональные вектора, единичной длины. λ_i показывает веса, с которыми надо взять эти вектора, пытаясь заменить каждый x_i. $k-$ исходная размерность. μ – середина облака. p – размерность, до которой хотим снизить. Для визуализации обычно берут $p = 2$.

Предположим, что V уже найдено. Оказывается, неважно, до какой размерности мы снижали, $\mu = \frac{\sum x_i}{n}$, а λ_i находится легко. Допустим μ, V найдены. Тогда

$$\lambda_1 : \min_{\lambda} {}_1||x_1 - \mu - V\lambda_1||^2$$

Это есть аналог МНК

$$\beta : \min_{\beta} ||y - X\beta||^2$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

В нашем случае:

$$\lambda_i = (V^T V)^{-1} V^T (x_i - \mu) = V^T (x_i - \mu)$$

На самом деле метод главных компонент - это выбор плоскости, а дальше МНК.

Почему $V^T V = I$?

$x_1, x_2, x_3, \ldots$ - вектора наблюдений. x_i - i-ое наблюдение, в отличие от МНК, где $x_1, x_2, x_3, \ldots$ - регрессоры. Ищем подпространство, чтобы при проецировании на него получались $\hat{x_1}, \hat{x_2}, \ldots$. Это подпространство будем задавать базисом $v_1, \ldots, v_p$. Для удобства задаем подпространство ортонормальным базисом.

$$\begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

λ_i мы уже нашли с помощью МНК, теперь осталось проминизировать по μ, V.

$$\min_{\mu,V} ||(x_i - \mu) - VV^T(x_i - \mu)||^2 = \min_{\mu,V} Q_\mu$$

VV^T - проектор, матрица-шляпница.

Выпишем условие первого порядка:

$$dQ_\mu = \sum_i ||(I - VV^T)(x_i - \mu)||^2 = \sum_i d((x_i - \mu)^T (I - VV^T)^T (I - VV^T)(x_i - \mu)) = \sum_i d(x_i - \mu)^T (I - VV^T)(x_i - \mu) =$$

$$= \sum_i -2(x_i - \mu)^T (I - VV^T) d\mu = 0$$

В итоге $\hat{\mu} = \frac{\sum x_i}{n}$ зануляет dQ. Но оно нединственно, можно придумать другие.