Мультиколлинеарность

5 ноября 2025

Строгая мультиколлеарность

Согласно одному из допущений Гаусса–Маркова, для того, чтобы модель линейной регрессии была идентифицируема (то есть, чтобы мы могли получить оценки коэффициентов), мы должны исключить случай строгой мультиколлеарности. Иными словами, не должно быть строгой линейной зависимости между предикторами. То есть, $rank(X) = p$, где $p = k + 1$ (количество параметров в регрессионной модели)

Если $rank(X) < p$, то $det(X^T X) = 0$ (вырожденная матрица $X^T X$). Следовательно, мы не можем получить оценки коэффициентов в линейной регрессии посредством МНК в условиях строгой мультиколлинеарности

Последствия нестрогой мультиколлинеарности

Мультиколлинеарность:

- Строгая
- Нестрогая (корреляция между предикторами)

Если корреляция между предикторами достаточно сильная (почти линейно зависимы), тогда $det(X^T X) \to 0$. Если мы даже немного поменяем массив данных, $\hat{\beta}$ будут меняться во много раз. Таким образом, последствия сильной мультиколлинеарности следующие:

1. неустойчивые результаты (оценки коэффициентов)
2. большие стандартные ошибки оценок коэффициентов, то есть, чаще будем делать вывод о незначимости коэффициентов (даже при условии высокого R^2)
3. при этом оценки коэффициентов остаются BLUE

Диагностики мультиколлинеарности:

1. визуализация (графики – попарные диаграммы рассеяния)
2. корреляционная матрица предикторов
3. коэффициент «вздутия» дисперсии (VIF: variance inflation factor)

$$Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2} VIF_j = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2(1 - R_j^2)}$$

R_j^2 – коэффициент детерминации вспомогательной регрессионной модели x_j на все остальные предикторы

$$VIF_j = \frac{1}{1 - R_j^2}$$

VIF больше 10 говорит о сильной мультиколлинеарности

Выведем формулу для $Var(\hat{\beta}_j)$

1. Пусть дана исходная модель:
   $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + ... + \hat{\beta}_k x_{ki}$
2. По теореме Фриша-Во-Ловелла получим $\hat{\beta}_j$: для этого очистим отклик от всех предикторов кроме j-го, а также очистим j-ый предиктор от всех остальных объясняющих переменных:
   $$\vec{\hat{y}_i} = X_{-j}\vec{\hat{\alpha}}$$
   Сохраняем $\hat{w}_i$ – остатки из этой модели (очищенный y)
   $$\vec{\hat{x}_j} = X_{-j}\vec{\hat{\gamma}}$$
   Сохраняем $\hat{u}_i$ – остатки из этой модели (очищенный x_j)
3. Оцениваем регрессию $\hat{w}_i$ на $\hat{u}_i$: $\widehat{\beta}_j = \dfrac{Cov(\hat{w}_i, \hat{u}_i)}{Var(\hat{u}_i)} = \dfrac{\hat{u}^T\hat{w}}{\hat{u}^T\hat{u}}$

$Var(\hat{\beta}_j)$ и VIF_j

$$\frac{\widehat{Cov}(\hat{w}_i, \hat{u}_i)}{\widehat{Var}(\hat{u}_i)} = \frac{\hat{u}^T\hat{w}}{\hat{u}^T\hat{u}} = \frac{\hat{u}^T(y_i - \hat{y}_{(-j)i})}{\hat{u}^T\hat{u}} = \frac{\hat{u}^T y}{\hat{u}^T\hat{u}}$$

$$Var(\hat{\beta}_j) = Var(\frac{\hat{u}^T y}{\hat{u}^T\hat{u}}) = \frac{\hat{u}^T\sigma^2 I\hat{u}}{(\hat{u}^T\hat{u})^2} = \frac{\sigma^2\hat{u}^T\hat{u}}{(\hat{u}^T\hat{u})^2} = \frac{\sigma^2}{\hat{u}^T\hat{u}}$$

$$\hat{u}^T\hat{u} = RSS_j = \sum_{i=1}^{n}(x_{ji} - \hat{x}_{ji})^2 = TSS_j(1-R_j^2) = \sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2(1-R_j^2)$$

$$VIF_j = \frac{1}{1-R_j^2}$$

показывает, во сколько раз увеличивается дисперсия $\hat{\beta}_j$ при мультиколлинеарности по сравнению со случаем, когда j-ый предиктор ортогонален другим предикторам

Что делать в случае сильной мультиколлинеарности?

1. убрать лишние предикторы (чтобы информация не дублировалась)
2. использовать метод главных компонент для снижения размерности признакового пространства
3. использовать методы регуляризации (к примеру, гребневую регрессию)
4. ничего не делать: как, к примеру, в случае моделей с переменными взаимодействия

Метод снижения размерности

Что имеем на «входе» задачи?

Ряд характеристик – признаков объектов, достаточно сильно скоррелированных между собой. Если информация дублируется, то почему бы не уменьшить количество признаков?

Метод снижения размерности

Что имеем на «входе» задачи?

Ряд характеристик – признаков объектов, достаточно сильно скоррелированных между собой. Если информация дублируется, то почему бы не уменьшить количество признаков?

Постановка задачи снижения размерности

От исходного признакового пространства размерности k, содержащего много дублирующей информации, перейти к новому признаковому пространству размерности p, где $p < k$. Корреляция между новыми признаками равна 0, то есть, дублирующая информация отсутствует.

Примеры: возможность построения индекса

Пример 1

1. количество раз обработки рук антисептиком за день
2. страх выходить на улицу
3. страх потери работы

Примеры: возможность построения индекса

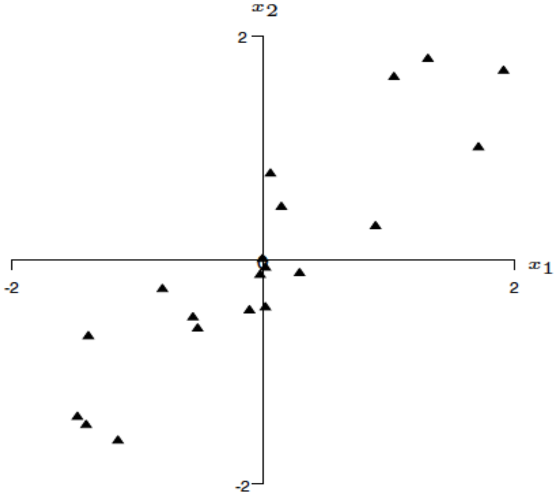
Пример 1

1. количество раз обработки рук антисептиком за день
2. страх выходить на улицу
3. страх потери работы

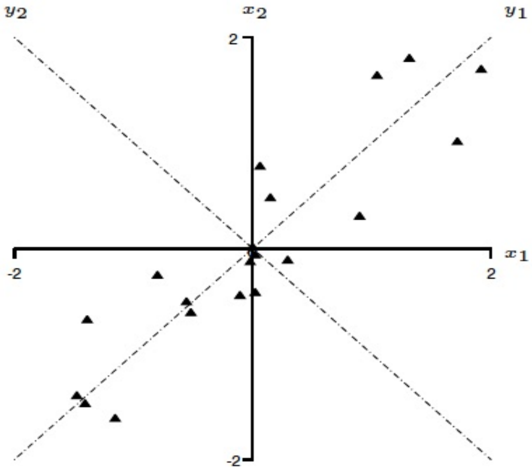
Пример 2

1. доверие правительству
2. доверие парламенту
3. доверие политическим партиям

Визуализация: на «входе» задачи



Визуализация: в результате преобразования



Как выполнить преобразование?

Алгоритм действий:

1. Переход к новой системе координат
   - Поворачиваем оси так, чтобы угол между осями составлял 90°.
   - Минимизация суммы квадратов расстояний от точки до прямой.
2. Оставляем оси с большей долей дисперсии (информации), остальные оси отбрасываем.

Переход к новой системе координат

$\overrightarrow{y} = A\overrightarrow{x}$ (главные компоненты – по строкам)

Для перехода от $\overrightarrow{x}$ к $\overrightarrow{y}$ нам понадобится ортогональная матрица A.

- как столбцы, так и строки имеют длину $= 1$
- как столбцы, так и строки ортогональны
- $|\det(\mathrm{A})| = 1$
- $A^{-1} = A^T$

Про длину и ортогональность

Пример

$$\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

- Длина (вектор по первой строке) : $\sqrt{\left(\frac{1}{\sqrt{5}}\right)^2 + \left(\frac{-2}{\sqrt{5}}\right)^2} = 1$
- Ортогональность первого и второго столбцов: $\frac{1}{\sqrt{5}}\left(\frac{-2}{\sqrt{5}}\right) + \frac{2}{\sqrt{5}\sqrt{5}} = 0$ (скалярное произведение $= 0$)

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$

Вариация новых показателей:

$\overrightarrow{y}^T\overrightarrow{y} = (A\overrightarrow{x})^T(A\overrightarrow{x}) = \overrightarrow{x}^T A^T A\overrightarrow{x} =$

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$

Вариация новых показателей:

$\overrightarrow{y}^T\overrightarrow{y} = (A\overrightarrow{x})^T(A\overrightarrow{x}) = \overrightarrow{x}^T A^T A\overrightarrow{x} = \overrightarrow{x}^T I\overrightarrow{x} = \overrightarrow{x}^T\overrightarrow{x}$

Что происходит с информацией?

Представим вариацию в матричном виде:

Вариация исходных показателей: $\overrightarrow{x}^T\overrightarrow{x}$
Вариация новых показателей:
$\overrightarrow{y}^T\overrightarrow{y} = (A\overrightarrow{x})^T(A\overrightarrow{x}) = \overrightarrow{x}^T A^T A\overrightarrow{x} = \overrightarrow{x}^T I\overrightarrow{x} = \overrightarrow{x}^T\overrightarrow{x}$

Принцип сохранения информации

Переходим от k-мерного к k-мерному ортогонализированному признаковому пространству (до удаления малоинформативных индексов). В результате МГК информация перераспределяется, при этом сохраняется полностью.

Для реализации МГК вспомним материал:

Собственные векторы и собственные значения

$M\overrightarrow{z} = \lambda\overrightarrow{z}$, где M – квадратная матрица, $\overrightarrow{z}$ – собственный вектор (ненулевой), λ – собственное значение.
В результате воздействия матрицы M получаем вектор, параллельный исходному вектору.

Для реализации МГК вспомним материал:

Собственные векторы и собственные значения

$M\overrightarrow{z} = \lambda\overrightarrow{z}$, где М – квадратная матрица, $\overrightarrow{z}$ – собственный вектор (ненулевой), λ – собственное значение.
В результате воздействия матрицы М получаем вектор, параллельный исходному вектору.

Характеристическое уравнение

$M\overrightarrow{z} = \lambda\overrightarrow{z} = \lambda I\overrightarrow{z}$, где I – единичная матрица.
$(M - \lambda I)\overrightarrow{z} = 0$
Может быть либо одно решение (нулевой $\overrightarrow{z}$), либо много решений. Для второго исхода матрица $(M - \lambda I)$ должна преобразовывать в 0 (вырожденная матрица):

$$\boxed{det(M - \lambda I) = 0}$$

Работа с матрицами в рамках МГК

Имеем: ковариационная/корреляционная матрица C

Матрица C содержит информацию об исходном признаковом пространстве. Вспомним общий вид ковариационной матрицы:

$$\begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

Работа с матрицами в рамках МГК

Имеем: ковариационная/корреляционная матрица C

Матрица C содержит информацию об исходном признаковом пространстве. Вспомним общий вид ковариационной матрицы:

$$\begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

Находим в процессе: матрицу преобразования A

Матрица A – матрица, составленная из собственных векторов матрицы C.

Работа с матрицами в рамках МГК

Имеем: ковариационная/корреляционная матрица C

Матрица C содержит информацию об исходном признаковом пространстве. Вспомним общий вид ковариационной матрицы:

$$\begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

Находим в процессе: матрицу преобразования A

Матрица A – матрица, составленная из собственных векторов матрицы C.

В результате: главные компоненты

Главная компонента (ГК): $y_{ji} = z_{1j}x_{1i} + \ldots z_{kj}x_{ki}$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda\overrightarrow{z}}$

Решим характеристическое уравнение:

$$\det\begin{pmatrix} 34-\lambda & 5 \\ 5 & 10-\lambda \end{pmatrix} = 0$$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda \overrightarrow{z}}$

Решим характеристическое уравнение:

$$\det \begin{pmatrix} 34 - \lambda & 5 \\ 5 & 10 - \lambda \end{pmatrix} = 0 \to (34 - \lambda)(10 - \lambda) - 25 = 0$$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda \overrightarrow{z}}$

Решим характеристическое уравнение:

$\det\begin{pmatrix} 34-\lambda & 5 \\ 5 & 10-\lambda \end{pmatrix} = 0 \rightarrow (34-\lambda)(10-\lambda) - 25 = 0$

$\rightarrow \lambda^2 - 44\lambda + 315 = 0$

Построим главные компоненты:

Пример:

Дана ковариационная матрица $C = \begin{pmatrix} 34 & 5 \\ 5 & 10 \end{pmatrix}$

Решение:

$\boxed{C\overrightarrow{z} = \lambda\overrightarrow{z}}$

Решим характеристическое уравнение:

$\det\begin{pmatrix} 34-\lambda & 5 \\ 5 & 10-\lambda \end{pmatrix} = 0 \to (34-\lambda)(10-\lambda) - 25 = 0$

$\to \lambda^2 - 44\lambda + 315 = 0 \to \lambda_1 = 35; \lambda_2 = 9$

$\lambda_1 = 35$ – вариация первой ГК

$\lambda_2 = 9$ – вариация второй (менее информативной) ГК

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$$(C - 35I)\overrightarrow{z} = 0$$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix}\overrightarrow{z} = 0 \rightarrow -z_1 + 5z_2 = 0 \rightarrow z_1 = 5z_2$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$.

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix}\overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = 5z_2 \\ 26z_2^2 = 1 \end{cases} \Rightarrow$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix}\overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = 5z_2 \\ 26z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{5}{\sqrt{26}} \\ z_2 = \dfrac{1}{\sqrt{26}} \end{cases}$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 35I)\overrightarrow{z} = 0$

$$\begin{pmatrix} -1 & 5 \\ 5 & -25 \end{pmatrix} \overrightarrow{z} = 0 \to -z_1 + 5z_2 = 0 \to z_1 = 5z_2$$

Для нахождения конкретных значений собственного вектора введем ограничение (исходя из свойств ортогональной матрицы): $z_1^2 + z_2^2 = 1$. Запишем как систему уравнений:

$$\begin{cases} z_1 = 5z_2 \\ z_1^2 + z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = 5z_2 \\ 26z_2^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{5}{\sqrt{26}} \\ z_2 = \dfrac{1}{\sqrt{26}} \end{cases}$$

Запишем ГК (1): $\boxed{\mathrm{y}_{1i} = \dfrac{5}{\sqrt{26}} x_{1i} + \dfrac{1}{\sqrt{26}} x_{2i}}$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0 \begin{pmatrix} 25 & 5 \\ 5 & 1 \end{pmatrix} \overrightarrow{z} = 0$

Запишем как систему уравнений:

$$\begin{cases} -5z_1 = z_2 \\ z_1^2 + 25z_1^2 = 1 \end{cases} \Rightarrow$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0 \begin{pmatrix} 25 & 5 \\ 5 & 1 \end{pmatrix} \overrightarrow{z} = 0$

Запишем как систему уравнений:

$$\begin{cases} -5z_1 = z_2 \\ z_1^2 + 25z_1^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{1}{\sqrt{26}} \\ z_2 = \dfrac{-5}{\sqrt{26}} \end{cases}$$

Построим главные компоненты:

Решение (продолжение):

Подставим последовательно собственные числа:

$(C - 9I)\overrightarrow{z} = 0 \begin{pmatrix} 25 & 5 \\ 5 & 1 \end{pmatrix} \overrightarrow{z} = 0$

Запишем как систему уравнений:

$$\begin{cases} -5z_1 = z_2 \\ z_1^2 + 25z_1^2 = 1 \end{cases} \Rightarrow \begin{cases} z_1 = \dfrac{1}{\sqrt{26}} \\ z_2 = \dfrac{-5}{\sqrt{26}} \end{cases}$$

Запишем ГК (2): $\boxed{\mathrm{y}_{2i} = \dfrac{1}{\sqrt{26}} x_{1i} + \dfrac{-5}{\sqrt{26}} x_{2i}}$

Регуляризация: метод гребневой регрессии

Гребневая регрессия накладывает штраф на большие по модулю коэффициенты:

$$\min_{\beta}\left(\mathrm{RSS} + \alpha\sum_{j=1}^{k}\beta_j^2\right)$$

Мы изменяем матрицу X^TX, добавляя константу α к её диагональным элементам

Получается, что

$$\hat{\beta}_{ridge} = (X^TX + \alpha I)^{-1}X^TY$$

, где α – это параметр регуляризации

- Матрица $X^TX + \alpha I$ обратима
- Получаем более устойчивые результаты
- Оценки коэффициентов получаются смещенными, однако уменьшается их дисперсия

Гребневая регрессия: поиск параметра α

k-блочная кросс-валидация:

1. массив разбивается на k равных подвыборок
2. для каждого заданного параметра α k итераций: j-ая подвыборка выступает тестовой, остальные подвыборки составляют обучающую
3. считаем среднее MSE по k итерациям для каждого значения α
4. выбираем то значение α, при котором усредненное MSE принимает минимальное значение