

# 1. Многомерный регрессионный анализ

## Основа алгоритмов байесовского оценивания - Теорема о многомерном условном распределении вероятностей:

Процесс в  $k$  векторов с НЗР, разделен на 2 части  $k_1$  (фактор, объясняющая часть) и  $k_2$  (отклик, результирующая часть) ( $k_1 + k_2 = k$ )  $\rightarrow$  описывается многомерным условным законом распределения вероятностей с оценками характеристик известного вида для условного математического ожидания  $MO(Y|X)$  и условной ковариационной матрицы  $K(Y|X)$ .

# 1. Многомерный регрессионный анализ

Основные шаги:

1. Расширенная матрица плана  $X_0$  совместная для  $k_1$  объясняющих переменных  $X$  и  $k_2$  результирующих переменных  $Y$  – они обязательно на последнем месте

$$X_0 = \left( \begin{array}{ccc|ccc} x_{11} & \boxtimes & x_{k_1 1} & y_{11} & \boxtimes & y_{k_2 1} \\ x_{12} & \boxtimes & x_{k_1 2} & y_{12} & \boxtimes & y_{k_2 2} \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1n} & \boxtimes & x_{k_1 n} & y_{1n} & \boxtimes & y_{k_2 n} \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes & \boxtimes \end{array} \right) = (X \quad Y)$$

$X_{nk_1}$ 
 $Y_{nk_2}$

# 1. Многомерный регрессионный анализ

2. Из матрицы плана  $X_0$  :

а) вектор средних по столбцам для  $X \rightarrow \bar{X}$  и  $Y \rightarrow \bar{Y}$

б) оценка совместной ковариационной матрицы  $K_0$ , которая состоит из блоков

$$K_0 = X_0^T \cdot X_0 / n = \frac{1}{n} \begin{pmatrix} X_{\zeta}^T X_{\zeta} & X_{\zeta}^T Y_{\zeta} \\ Y_{\zeta}^T X_{\zeta} & Y_{\zeta}^T Y_{\zeta} \end{pmatrix} = \begin{pmatrix} K_{XX} & K_{XY} \\ K_{YX} & K_{YY} \end{pmatrix}$$

где  $X_{\zeta}$ ,  $Y_{\zeta}$  – вектора, центрированные средним, для получения ковариационной матрицы. Использование девиаций  $S_{ij}$  (не нормированные величиной  $n$  отклонения от среднего как  $S$  вместо  $K_0$  ).

# 1. Многомерный регрессионный анализ

3. Характеристики многомерного условного распределения вероятностей:

– условное математическое ожидание (линейное уравнение регрессии с многомерным откликом)

$$MO(Y | X) = \hat{Y} = \bar{Y} + K_{YX} \cdot K_{XX}^{-1} (X - \bar{X})$$

$k_2 \times 1$        $k_2 \times 1$        $k_2 \times 1$        $k_2 \times k_1$        $k_1 \times k_1$        $k_1 \times 1$

с соответствующими размерностями, т.к.

$$K_0 = \begin{pmatrix} K_{XX} & K_{XY} \\ K_{YX} & K_{YY} \end{pmatrix}, \quad k = k_1 + k_2$$

$k_1 \times k_1$        $k_1 \times k_2$   
 $k_2 \times k_1$        $k_2 \times k_2$

Очевидно что вектора средних – столбцы!

# 1. Многомерный регрессионный анализ

Обычно условное математическое ожидание (линейное уравнение регрессии с многомерным откликом) приводят к нормальному виду

$$\begin{aligned} MO(Y | X) = \hat{Y} &= \bar{Y} + K_{YX} \cdot K_{XX}^{-1} (X - \bar{X}) = \bar{Y} + A(X - \bar{X}) = \\ &= \bar{Y} + AX - A\bar{X} \end{aligned}$$

где

$$A = K_{YX} \cdot K_{XX}^{-1}$$

Тогда окончательно его нормальный вид с размерами

$$\underset{k_2 \times 1}{\hat{Y}} = AX + (\bar{Y} - A\bar{X}) = \underset{k_2 \times k_1}{A} \cdot \underset{k_1 \times 1}{X} + \underset{k_2 \times 1}{B}$$

с вектором свободных членов  $B = \bar{Y} - A\bar{X}$

# 1. Многомерный регрессионный анализ

– условная ковариационная матрица  $K(Y|X)$  отклика (результатирующей матрицы переменных  $Y$  модели) с размерностями

$$K_{Y|X} = K_{YY} - K_{YX} \cdot K_{XX}^{-1} \cdot K_{XY}.$$

$k_2 \times k_2$        $k_2 \times k_2$        $k_2 \times k_1$        $k_1 \times k_1$        $k_1 \times k_2$

Здесь  $K_{Y|X}$  - матрица оценок точности смоделированных рядов откликов  $\hat{Y}$  – по диагонали – дисперсии, недиагональные – ковариации.

Формулы позволяют решить задачу определения оптимальных коэффициентов линейного преобразования одной части в другую с оценкой точности модели преобразования.

# 1. Многомерный регрессионный анализ

Вариант когда известна обратная ковариационная матрица

$K_0^{-1} = C_0$ . 2 основных подхода:

1. Обратить обратную матрицу и использовать полученные ранее формулы;

2. Воспользоваться теоремой Фробениуса об обращении блочных матриц

$$K_0 = \begin{pmatrix} K_{XX} & K_{XY} \\ K_{YX} & K_{YY} \end{pmatrix} \rightarrow K_0^{-1} = C_0 = \begin{pmatrix} K_{XX}^{-1} + U^T M^{-1} U & -U^T M^{-1} \\ -M^{-1} U & M^{-1} \end{pmatrix} =$$

$$= \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix}$$

Здесь

$$M = K_{YY} - K_{YX} K_{XX}^{-1} K_{XY}, \quad U = K_{YX} K_{XX}^{-1}$$

# 1. Многомерный регрессионный анализ

Вспоминаем, что матрица коэффициентов  $A$  и матрица оценок  $K_{Y|X}$  имеют вид

$$A = K_{YX} \cdot K_{XX}^{-1}$$

$$K_{Y|X} = K_{YY} - K_{YX} \cdot K_{XX}^{-1} \cdot K_{XY}$$

Рассматривая структуру обратной матрицы  $C_0$

$$C_0 = \begin{pmatrix} K_{XX}^{-1} + U^T M^{-1} U & -U^T M^{-1} \\ -M^{-1} U & M^{-1} \end{pmatrix} = \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix}$$

$$M = K_{YY} - K_{YX} K_{XX}^{-1} K_{XY}, \quad U = K_{YX} K_{XX}^{-1}$$

не сложно заметить, что для коэффициентов в виде «строки» имеем

а в виде «столбца» соответственно

$$A = U^T = U^T M^{-1} \cdot M = -C_{XY} \cdot C_{YY}^{-1}$$

# 1. Многомерный регрессионный анализ

Матрица оценок  $K_{Y|X}$

$$K_{Y|X} = K_{YY} - K_{YX} \cdot K_{XX}^{-1} \cdot K_{XY}$$

из структуры обратной матрицы  $C_0$

$$C_0 = \begin{pmatrix} K_{XX}^{-1} + U^T M^{-1} U & -U^T M^{-1} \\ -M^{-1} U & M^{-1} \end{pmatrix} = \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix}$$

$$M = K_{YY} - K_{YX} K_{XX}^{-1} K_{XY}, \quad U = K_{YX} K_{XX}^{-1}$$

есть  $K_{Y|X} = M = C_{YY}^{-1}$

Нюанс нормировки ( $n$  и  $n - k$ ). Целесообразность девиационной матрицы  $S$ . Формулы этого вида используются часто.

# 1. Многомерный регрессионный анализ

Формулы получают исходя из следующих соображений:

- Для всего процесса с  $k$  рядами получают многомерный закон распределения  $f(Y, X)$  (совместный для набора  $X$  из  $k_1$  рядов и  $Y$  из  $k_2$  рядов)
- Получаем многомерный закон распределения  $f(X)$  для набора  $X$  из  $k_1$  факторных переменных.
- По теореме Байеса условный закон распределения  $f(Y|X)$  для  $Y$  при фиксированных (измеренных) рядах  $X$  получаем как

$$f(Y | X) = \frac{f(Y, X)}{f(X)}$$

Новый закон распределения  $f(Y|X)$  имеет главные характеристики: условное математическое ожидание  $MO(Y|X)$  и условную ковариационную матрицу  $K_{Y|X}$

# 1. Многомерный регрессионный анализ

## Основные частные случаи теоремы:

1. 1 факторная переменная (ряд), 1 результирующая – парный линейный регрессионный анализ
2. Много факторных переменных, 1 результирующая – многомерный (многофакторный) линейный регрессионный анализ с одномерным откликом (1-откликом)
3. Много факторных переменных, много результирующих – многомерный (многофакторный) линейный регрессионный анализ с многомерным откликом ( $n$ -откликом)

Известная и важная в геодезии задача трансформации систем координат имеет:

2 факторные, 2 результирующие переменные – 2-факторный линейный регрессионный анализ с 2-откликом.

Расчет по девиатам. Девиационная матрица. Коэффициенты – по условному математическому ожиданию, целевую функцию  $v^T v$  – по условной ковариационной матрице

# 1. Многомерный регрессионный анализ

1 вариант: – 1 фактор, 1 отклик:

Характеристики одномерного условного закона распределения для 1-отклика  $MO(y|x) = f(x)$  для процесса с факторной переменной  $x$  и результирующей переменной  $y$  на последнем месте ( $x y$ ): имеется выборочная ковариационная матрица

$$K_0 = \begin{pmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{pmatrix} = \begin{pmatrix} D_x & \text{cov}(x, y) \\ \text{cov}(x, y) & D_y \end{pmatrix} \text{ - элементы числа,}$$

Понадобится обратная ковариационная матрица  $C_0$

$$K_0^{-1} = C_0 = \begin{pmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{pmatrix} = \frac{1}{(1 - r_{xy}^2)} \begin{pmatrix} \frac{1}{D_x} & \frac{-r_{xy}}{\sigma_x \sigma_y} \\ \frac{-r_{xy}}{\sigma_x \sigma_y} & \frac{1}{D_y} \end{pmatrix}$$

# 1. Многомерный регрессионный анализ

– условное математическое ожидание ( и она же линейная форма парной регрессии)

$$\begin{aligned} MO(y|x) &= \hat{y} = \bar{y} + k_{yx} \cdot k_{xx}^{-1} (x - \bar{x}) = \\ &= \bar{y} + \frac{k_{yx}}{k_{xx}} \cdot (x - \bar{x}) = \bar{y} + \frac{\text{cov}(x, y)}{D_x} \cdot (x - \bar{x}) = \bar{y} + r_{xy} \cdot \frac{\sigma_y}{\sigma_x} \cdot (x - \bar{x}) \\ &= \bar{y} + a' (x - \bar{x}) \end{aligned}$$

или в нормальном виде

$$\hat{y} = \bar{y} + a' (x - \bar{x}) = a' \cdot x + (\bar{y} - a' \bar{x}) = a' \cdot x + d$$

с

$$a' = \frac{\sigma_y}{\sigma_x} \cdot r_{xy} = \frac{\text{cov}(y, x)}{D_x}, \quad d = \bar{y} - a' \cdot \bar{x}$$

# 1. Многомерный регрессионный анализ

– условная дисперсия (ковариационная матрица для 1 ряда, дисперсия модели)

$$\begin{aligned} K_{y|x} = D(y|x) &= \hat{\alpha}_0^2 = k_{yy} - k_{yx} \cdot k_{xx}^{-1} \cdot k_{xy} = D_y - \frac{\text{cov}^2(x, y)}{D_x} = \\ &= D_y - \frac{r_{xy}^2 \cdot D_x \cdot D_y}{D_x} = D_y \cdot (1 - r_{xy}^2) \end{aligned}$$

Проблема нормировки: что получим умножают на

$$t = \frac{n}{n - 2}$$

т.к. два определяемых коэффициента. Если матрица девиат  $S$ , то что получают (это  $[v^2]$ ) делят на  $(n - 2)$ .

# 1. Многомерный регрессионный анализ

Вычисления через прямую  $K_0$  и обратную  $C_0$   
ковариационные матрицы:

- прямая

$$K_0 = \begin{pmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{pmatrix} = \begin{pmatrix} D_x & \text{cov}(x, y) \\ \text{cov}(x, y) & D_y \end{pmatrix}$$

коэффициенты регрессии

$$a' = k_{yx} \cdot k_{xx}^{-1} = \frac{k_{yx}}{k_{xx}}, \quad d = \bar{y} - a' \cdot \bar{x}$$

дисперсия модели

$$D(y|x) = \sigma_0^2 = k_{yy} - k_{yx} \cdot k_{xx}^{-1} \cdot k_{xy} = k_{yy} - \frac{k_{xy}^2}{k_{xx}} = \frac{\det(K_0)}{k_{xx}}$$

Не забыть нормировку t.

# 1. Многомерный регрессионный анализ

- обратная

$$K_0^{-1} = C_0 = \begin{pmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{pmatrix} = \frac{1}{(1 - r_{xy}^2)} \begin{pmatrix} \frac{1}{D_x} & \frac{-r_{xy}}{\sigma_x \sigma_y} \\ \frac{-r_{xy}}{\sigma_x \sigma_y} & \frac{1}{D_y} \end{pmatrix}$$

коэффициенты регрессии

$$a' = -c_{xy} \cdot c_{yy}^{-1} = -\frac{c_{xy}}{c_{yy}}, \quad d = \bar{y} - a' \cdot \bar{x}$$

дисперсия модели

$$D(y|x) = \hat{\sigma}_0^2 = c_{yy}^{-1}$$

Не забыть нормировку  $t$ . Трудоемка точность коэффициентов.

# 1. Многомерный регрессионный анализ

2 вариант: 1 отклик,  $n$  факторов

Теорема о характеристиках многомерного условного закона распределения для 1-отклика: имеется выборочная ковариационная матрица

$$K_0 = \begin{pmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{pmatrix} = \begin{pmatrix} K_{xx} & k_{xy} \\ k_{yx} & k_y \end{pmatrix}$$

для процесса с факторными переменными  $X = (x_1, x_2, \dots, x_k)$  и результирующей переменной  $y$  на последнем месте ( $X \ y$ ). Закон распределения процесса описывается условным многомерным нормальным законом распределения вероятностей характеристиками:

# 1. Многомерный регрессионный анализ

– условным математическим ожиданием (линейной формой множественной регрессии)

$$MO(y | x_1, \dots, x_n) = \hat{y} = \bar{y} + k_{yx} \cdot K_{xx}^{-1} (x - \bar{x}) = \bar{y} + a' (x - \bar{x})$$

или в нормальном виде

$$\hat{y} = \bar{y} + a' (x - \bar{x}) = a' \cdot x + (\bar{y} - a\bar{x}) = a' \cdot x + d$$

Здесь  $a'$  – вектор-строка,  $x$  и  $y$  – столбцы.

– условной дисперсией (дисперсией модели)

$$D(y | x_1, \dots, x_n) = \hat{\sigma}_0^2 = k_y - k_{yx} \cdot K_{xx}^{-1} \cdot k_{xy}$$

или

$$\hat{\sigma}_0^2 = D_y (1 - r_{y|\dots}^2)$$

# 1. Многомерный регрессионный анализ

Через обратную матрицу  $C_0$

$$C_0 = \begin{pmatrix} K_{XX}^{-1} + U^T M^{-1} U & -U^T M^{-1} \\ -M^{-1} U & M^{-1} \end{pmatrix} = \begin{pmatrix} K_{XX}^{-1} + \frac{u^T \cdot u}{m} & -\frac{u^T}{m} \\ -\frac{u}{m} & \frac{1}{m} \end{pmatrix} =$$
$$= \begin{pmatrix} C_{XX} & c_{Xy} \\ c_{yX} & c_y \end{pmatrix} \quad m = k_y - k_{yX} K_{XX}^{-1} k_{Xy}, \quad u = k_{yX} K_{XX}^{-1}$$

имеем вектор коэффициентов в виде строки и столбца

$$u = a' = \frac{c_{yX}}{c_y}, \quad u^T = a'^T = \frac{c_{Xy}}{c_y}$$

и дисперсию модели

$$K_{Y|X} = M = C_{YY}^{-1} = D_{y|X} = \sigma_0^2 = \frac{1}{c_y}$$

# 1. Многомерный регрессионный анализ

3 случай общий:  $n$  факторов,  $n$  откликов уже рассмотрен  
Нюанс для оценки модели:

$$K_{Y|X} = K_{YY} - K_{YX} \cdot K_{XX}^{-1} \cdot K_{XY} = K_{YY} - A \cdot K_{XY}$$

Результат – матрица, где по диагонали дисперсии смоделированных  $k_2$  рядов  $Y$ . Общая дисперсия – их сумма, т.е. *след* матрицы  $K_{Y|X}$ . Тогда погрешность модели в общем (среднем) есть

$$\sigma_0^2 = Tr(K_{Y|X})$$

Соблюдать нормировку, извлекать корень.