А. В. Ж а р к о в (Ульяновск, УлГУ)**. Применение нестатистического факторного анализа в задачах обработки данных.**

Предположим, что имеется матрица первичных данных $\Omega = (\omega_{ij})_{n\times m}$, где ω_{ij} (i-й объект, j-я характеристика объекта) — некоторое число, необязательно обладающее статистическими свойствами. Сформируем две факторные матрицы $K = \Omega^T\Omega$, $F = \Omega\Omega^T$. Справедливо следующее утверждение [I]: матрица может быть представлена в виде

$$\Omega = \sqrt{\lambda_{(1)}}f^{(1)}k^{(1)T} + \cdots + \sqrt{\lambda_{(r)}}f^{(r)}k^{(r)T}. \tag{1}$$

Здесь $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > 0$ — все положительные собственные числа матриц K и F, $\{(k^{(1)}, \ldots, k^{(m)}\}$ и $\{f^{(1)}, \ldots, f^{(n)}\}$ — их соответствующие единичные собственные векторы. При некоторых неограничительных предположениях разложение (1) единственно. Известно, что первое слагаемое в (1) представляет собой наилучшее (в смысле евклидовой нормы матриц) одноранговое приближение матрицы Ω.

Спектральное разложение (1) произвольной матрицы Ω может применяться (при некоторых разумных предположениях) в различных прикладных исследованиях в области педагогики [2], биологии, социальных науках. Но и некоторые известные задачи статистического анализа данных решаются с использованием разложения (1).

Рассмотрим задачу наилучшего линейного приближения таблично заданной функции, обычно решаемую методом наименьших квадратов (МНК). Пусть дана таблица значений некоторой функции (x_i, y_i), $i = 1, 2, \ldots, n$, причем значения функции $y(x)$ в таблице связаны с аргументом x линейным соотношением с некоторой «ошибкой»: $y_i = kx_i + b + \varepsilon_i$, $i = 1, 2, \ldots, n$. В МНК требуется, чтобы величина $S = \sum_{i=1}^{n} \varepsilon_i^2$ принимала минимальное значение при всех возможных k и b. Определим величины $\widetilde{x}_i = x_i - \overline{x}$, $\widetilde{y}_i = y_i - \overline{y}$, $i = 1, 2, \ldots, n$, и векторы $\widetilde{X} = (\widetilde{x}_i) \in \mathbf{R}^n$, $\widetilde{Y} = (\widetilde{y}_i) \in \mathbf{R}^n$, $E = (\varepsilon_i) \in \mathbf{R}^n$. Тогда $\widetilde{X} = k^{-1}(\widetilde{Y} - E)$. Введем матрицу данных $\Omega = (\widetilde{X}, \widetilde{Y})_{n\times 2}$.

Если применить к матрице Ω подход факторного анализа, то появится матрица $K = \Omega^T\Omega_{2\times 2}$, причем в случае статистического подхода $K = nM_2$, где M_2 — матрица выборочных вторых центральных моментов, подсчитанных по столбцам X и Y. Рассмотрим теперь сингулярное разложение (1) матрицы Ω. Первое слагаемое здесь можно записать в виде $\sqrt{\lambda_1}k_2^{(1)}f^{(1)}(k_1^{(1)}/k_2^{(1)}, 1)$, тогда получим линейную зависимость

$$y_i = \widehat{k}x_i + \widehat{b} + \widehat{\varepsilon}_i, \qquad \widehat{k} = k_2^{(1)}/k_1^{(1)}, \quad E = \widetilde{Y} - k_2^{(1)}\sqrt{\lambda_1}f^{(1)}, \tag{2}$$

аналогичную МНК. При этом, так как первое слагаемое в (1) является наилучшим одноранговым приближением Ω (в смысле евклидовой нормы разности), то величина нормы «ошибки» заведомо меньше величины S, полученной в МНК. Прямая (2) отличается от прямой МНК тем, что для нее «ошибки» измеряются перпендикулярно направляющему вектору прямой, а не вдоль оси Oy, как для модели МНК.

СПИСОК ЛИТЕРАТУРЫ

1. *Хорн Р., Джонсон Ч.* Матричный анализ. М.: Мир, 1989, 656 с.
2. *Жарков А. В.* Нестатистический факторный анализ в задаче оценивания успешности обучения. — Обозрение прикл. и промышл. матем., 2006, т. 14, в. 1, с. 110–111.