УДК 528.854.2

Сравнение плоских и многомерных регрессионных методов анализа данных

С.И. Суханов, Е.П. Крупочкин
АлтГУ, г. Барнаул

Регрессия представляет собой подход, с помощью которого можно соотнести друг с другом два набора переменных. Регрессионные модели, основаны на том, что сначала строится модель (X,Y) с помощью исходных данных (рисунок 1), затем построенная модель используется для предсказания (рисунок 2).

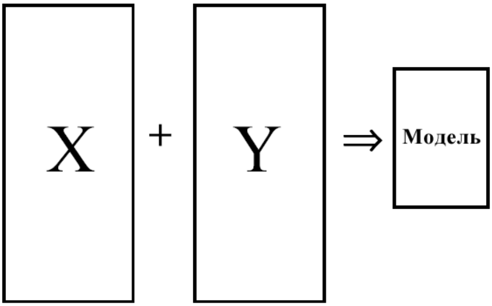

Рисунок 1 –Построение регрессионной модели по исходным данным X и Y

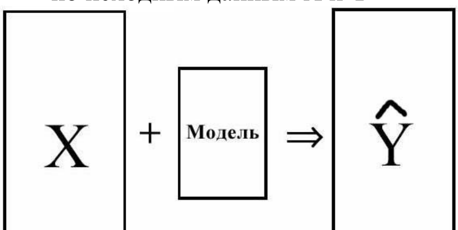

Рисунок 2 – Использование регрессионной модели для предсказания новых значений Y

Существует большое количество проекционных регрессионных методов, в частности: регрессия на главные компоненты (РГК), проекция на латентные структуры (ПЛС), многомерная регрессия (N-ПЛС).

Метод РГК можно рассматривать, как двухэтапную процедуру: сначала с помощью МГК преобразуется матрица X , а затем получившаяся матрица T (матрица счетов) непосредственно используется в модели: $Y = TB + E$ **,** где B – вектор регрессионных коэффициентов. Главным недостатком метода является то, что значения откликов Y при построении модели ни как не учитываются.

Для устранения этого недостатка существует проекция на латентные структуры (ПЛС). ПЛС представляет очень хорошую возможность интерпретации результатов. Он использует структуру данных Y и дисперсию Y напрямую, как некоторое руководство для декомпозиции матрицы X . В результате получаются оптимальные регрессионные соотношения, более точные с точки зрения процедуры проверки пригодности модели для прогнозирования, подробное описание данного метода изложено в [1].

Регрессии с помощью многомерных методов так же позволяет соотнести друг с другом два набора переменных. Одним из таких видов регрессии является N-ПЛС-регрессия. Главным её отличаем от ПЛС-регрессии является то, что матрица входных переменных не преобразуется в плоскую. А набор откликов Y теперь уже представляет собой, не вектор, а матрицу. N-ПЛС-регрессию можно рассматривать, как двухэтапную процедуру: сначала входная матрица X и выходная – Y подвергаются PARAFAC-разложению: $X = T \otimes W_j \otimes W_k$, $Y = U \otimes Q$ где $\otimes$ – кронекерово произведение. Матрица предсказаний Y^* находится следующим образом: новая входная матрица X^* также подвергается PARAFAC-разложению: $\mathrm{X}^* = \mathrm{T}^* \otimes \mathrm{W}_\mathrm{j}^* \otimes \mathrm{W}_\mathrm{k}^*$, $\mathrm{U}^* = \mathrm{T}^*\mathrm{B}$, где B – вектор регрессионных коэффициентов, матрица предсказаний Y^* строится с помощью кронекерового произведения двух матриц U^* и Q : $Y^* = U^* \otimes Q$., подробное описание данного метода изложено в [2].

Любую регрессию можно использовать в качестве классификатора, при этом в процессе разработки системы классификации необходимо ответить на следующие вопросы:

1. Как построить классификатор? Задача построения классификатора – выбор решающего правила, по которому на основании вектора признаков осуществляется отнесение объекта к тому или иному классу.
2. Как оценить качество построенной системы классификации? Задача количественной оценки системы (выбранные признаки + классификатор) с точки зрения правильности или ошибочности классификации.

Библиографический список

1. Эсбенсен, К. Анализ многомерных данных: избранные главы / К. Эсбенсен; Пер. с англ. С.В. Кучерявского; под ред. О.Е. Родионовой. –Барнаул: Изд-во Алт. гос. ун-та, 2003. – 150 с.
2. Bro, R. Multi–way analysis with applications in the chemical sciences / A. Smilde, B. Rasmus, P. Geladi. – Chichester: Wiley, 2004, 381 p.

УДК 517.95

Автомодельная задача фильтрации вязкой жидкости в вязкоупругом пористом скелете

М.А. Токарева, В.Н. Ларионова

АлтГУ, г. Барнаул

Рассмотрим следущую квазилинейную систему составного типа, описывающую пространственное нестационарное изотермическое движение сжимаемой жидкости в вязкоупругой пористой среде [1] – [2]:

$$\frac{\partial(1-\phi)\rho_s}{\partial t}+div((1-\phi)\rho_s\vec{v}_s)=0,$$
$$\frac{\partial(\rho_f\phi)}{\partial t}+div(\rho_f\phi\vec{v}_f)=0,$$
$$\phi(\vec{v}_f-\vec{v}_s)=-\frac{k\phi^n}{\mu}(\nabla p_f-\rho_f\vec{g}),$$
$$\nabla\cdot\vec{v}_s=-\frac{\phi^m}{\eta}p_e-\phi^b\beta_\phi(\frac{\partial p_e}{\partial t}+\vec{v}_s\cdot\nabla p_e),$$
$$\nabla\cdot\sigma+\rho_{tot}\vec{g}=0,\ \rho_{tot}=\phi\rho_f+(1-\phi)\rho_s,$$
$$p_{tot}=\phi p_v+(1-\phi)p_s, p_e=(1-\phi)\left(p_s-p_f\right).$$

Здесь ϕ – пористость; $\rho_f,\rho_s,\vec{v}_s,\vec{v}_f$ – соответственно истинные плотности и скорости фаз; p_e – эффективное давление, p_{tot} – общее давление, ρ_{tot} – общая плотность; $\vec{g}$ – плотность массовых сил; β_ϕ – коэффициент сжимаемости твердого скелета, η – динамическая вязкость твердой фазы, k – проницаемость, $\mu -$ динамическая вязкость жикости, σ – общий тензор напряжений. Истинная плотность твердой фазы ρ_s принимается постоянной. Система является замкнутой, если $p_f=p_f(\rho_f)$ или $\rho_f=const$. В случае неполного уравнения баланса сил $\frac{dp_{tot}}{dx_i}=-\rho_{tot}g$ и g=0 разрешимость автомодельной задачи для исходной системы уравнений установлена в работе [3].