А. В. Б е р н ш т е й н (Москва, ИСА РАН). **Усовершенствованные процедуры искусственных нейронных сетей.**

Известны наилучшие линейные процедуры в задачах построения зависимости между данными (метод наименьших квадратов) и снижения размерности множества многомерных данных (метод главных компонент). Среди нелинейных процедур (с помощью ядерных функций, радиальных базисных функций и др.) часто используются искусственные нейронные сети (ИНС), включенные в стандартные статистические пакеты. Однако эти процедуры являются неэффективными и для них могут быть построены мажорирующие процедуры.

Задача построения зависимости между переменными $x \in \mathbf{R}^p$ и $y \in \mathbf{R}^q$ по множеству значений $\{(x_i, y_i),\ i = 1, 2, \ldots, N\}$ состоит в построении функции $y = f_N(x)$, минимизирующей среднеквадратическую ошибку аппроксимации

$$\varepsilon = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |f_N(x_i) - y_i|^2}. \tag{1}$$

Оценка $f_{\text{ИНС},N}(x)$, построенная при помощи процедуры ИНС, заключается в последовательном применении трех преобразований:

$$x \in \mathbf{R}^p \ \to\ u = A_1 x \in \mathbf{R}^k \ \to\ v = F(u) \in \mathbf{R}^k \ \to\ y = A_2 v \in \mathbf{R}^q,$$

определяемых матрицами A_1 и A_2 размеров $k \times p$ и $q \times k$ соответственно, и функцией F, которая преобразует j-ю компоненту u_j вектора u в j-ю компоненту $v_j = \varphi(u_j)$ вектора v при помощи одной и той же одномерной функции φ. Тем самым, процедура $f_{\text{ИНС},N}(x)$ определяется набором параметров (k, φ, A_1, A_2). В качестве функции φ обычно априори выбирается конкретная сигмоидная функция (гиперболический косинус, гиперболический тангенс и т. п.). При фиксированном значении параметра k (размера «скрытого» слоя ИНС, оптимизация по которому проводится путем перебора), матрицы A_1 и A_2 выбираются из условия минимизации ошибки ε (1), основываясь на идеях обучения «с учителем» и кросс-проверки и используя итеративный алгоритм обратного распространения ошибок.

При фиксированном значении матрицы A_1 наилучшая матрица A_2 является решением задачи построения линейной регрессионной зависимости между переменными $F(A_1 x_i)$ и y_i, $i = 1, 2, \ldots, N$, и может быть выписана в явном виде. Использование этого факта позволяет построить процедуру минимизации, снижающую ошибку аппроксимации и время обучения в несколько раз.

Задача снижения размерности состоит в построении по множеству данных $\{x_i \in \mathbf{R}^p,\ i = 1, 2, \ldots, N\}$ процедуры $\Sigma = \{m, C_m, R_m\}$, определяемой размерностью m сжатых данных и преобразованиями сжатия C_m: $x \in \mathbf{R}^p \to \lambda = C_m(x) \in \mathbf{R}^m$ и восстановления R^m: $\lambda \in \mathbf{R}^m \to x = R_m(\lambda) \in \mathbf{R}^p$, минимизирующей меру δ близости между исходными и восстановленными векторами:

$$\delta = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \|x_i - R_m(C_m(x_i))\|^2}. \tag{2}$$

Репликативные ИНС (РИНС) в задаче снижения размерности определяются двумя цепочками преобразований:

$$\begin{aligned} C_{\text{ИНС},m}:\ & x \in \mathbf{R}^p \ \to\ u_1 = A_1 x \in \mathbf{R}^s \ \to\ v_1 = F_1(u_1) \in \mathbf{R}^s \ \to\ \lambda = A_2 v_1 \in \mathbf{R}^m, \\ R_{\text{ИНС},m}:\ & \lambda \in \mathbf{R}^m \ \to\ v_2 = A_3 \lambda \in \mathbf{R}^t \ \to\ u_2 = F_2(v_2) \in \mathbf{R}^t \ \to\ x = A_4 u_2 \in \mathbf{R}^p, \end{aligned}$$

и зависят от априори выбираемых сигмоидальных функций F_1 и F_2, размерностей s и t скрытых слоев и матриц A_i, $i = 1, 2, 3, 4$. Минимизация δ (2) проводится по этим матрицам при фиксированных значениях s и t, по которым затем проводится перебор. Можно показать, что в оптимальной ИРНС размерности s и t должны быть равны, а матрица A_1 явно определяется по матрице A_4 и равна $((A_4)^T A_4)^{-1}(A_4)^T$. Использование этого факта позволяет построить РИНС с меньшими ошибкой δ и временем обучения.