

АЛГОРИТМ ОБРАТНОГО РАСПРОСТРАНЕНИЯ ОШИБКИ ПРИ ОБУЧЕНИИ НЕЙРОСЕТИ ТИПА МНОГОСЛОЙНЫЙ ПЕРСЕПТРОН С ФУНКЦИЯМИ АКТИВАЦИИ ТИПА ГИПЕРБОЛИЧЕСКИЙ ТАНГЕНС

1. Инициализировать все веса $w_{ij}^{(k)}$ и смещения $w_{0j}^{(k)}$ нейронов случайными значениями в диапазоне $[-1, 1]$, где $i = \overline{1, N_{prev}}$ – номер нейрона предыдущего слоя, N_k – число нейронов в k -м слое, $j = \overline{1, N_{curr}}$ – номер нейрона текущего слоя, $k = \overline{2, M}$ (первый слой не имеет весов и не участвует в обучении), M – число слоев в нейросети.
2. Выбрать эталон $\bar{X}_i$ из множества эталонов $\mathfrak{R}$ для обучения, где $i = \overline{1, K}$ – номер эталона, $K = |\mathfrak{R}|$ – количество эталонов.
3. Подать данные выбранного эталона на вход нейросети. Начиная от слоев с меньшими номерами и переходя к слоям с большими номерами вычислять значения выходов нейронов:

$$Y_j^{curr} = f \left(w_{0j}^{curr} + \sum_{i=1}^{N_{prev}} w_{ij}^{curr} Y_i^{prev} \right),$$

где $Y_j^{(k)}$ – значение выхода j -го нейрона k -го слоя, $j = \overline{1, N_{curr}}$, $k = \overline{1, M}$, $f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$ – нелинейная функция (гиперболический тангенс).

4. Для каждого нейрона выходного слоя рассчитать значение ошибки:

$$\delta_i^{(k)} = (D_i - Y_i) \frac{df(x)}{dx} \Big|_{x=Y_i} = (D_i - Y_i)(1 - Y_i^2),$$

где D_i – требуемое значение выхода, $i = \overline{1, N_k}$. Скорректировать значения весов нейронов выходного слоя:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij},$$

$$\Delta w_{ij} = \eta \delta_j Y_i^{prev},$$

где t – номер итерации обучения, η – параметр скорости обучения, $i = \overline{1, N_{prev}}$, $j = \overline{1, N_{curr}}$.

5. Начиная со слоев с большими номерами и переходя к слоям с меньшими номерами вычислять значения ошибки и корректировать веса нейронов:

$$\delta_i^{curr} = \left(1 - (Y_i^{curr})^2 \right) \sum_{j=1}^{N_{next}} \delta_j^{next} w_{ij}^{next}, \quad i = \overline{1, N_{curr}}, \quad j = \overline{1, N_{next}},$$

$$w_{ij}^{curr}(t+1) = w_{ij}^{curr}(t) + \Delta w_{ij}^{curr},$$

$$\Delta w_{ij}^{curr} = \eta \delta_j^{curr} Y_i^{prev}, \quad i = \overline{1, N_{prev}}, \quad j = \overline{1, N_{curr}}.$$

6. Повторение шагов 2 – 6 до завершения обучения.

Диапазоны изменения значений признаков с составе эталонов необходимо нормировать.

Критерий окончания обучения может быть выражен в следующем виде:

$$(\eta < \eta_0) \vee \left(\max_{i=\overline{1, K}} \varepsilon_i < \varepsilon_0 \right),$$

где η_0 – пороговое значение параметра скорости обучения, ε_0 – пороговое значение ошибки (максимальная ошибка одного из эталонов $\varepsilon_i = \|\bar{Y}_i - \bar{D}_i\|$ меньше порога, $\|x\|$ – норма вектора, например, Евклидова).

Выбор очередного эталона $\bar{X}_j$ для обучения из множества эталонов следует производить пропорционально их ошибкам: $j = \arg\left(\max_{i=1, \bar{K}} \varepsilon_i \cdot r_i\right)$, где r_i – случайная величина с равномерным законом распределения, что уменьшает число итераций обучения (эталон, дающий большую ошибку, имеет большую вероятность быть показанным сети в процессе обучения).