

## 6. Метод обратного распространения точности для оценки среднеквадратических отклонений

Рассмотрим обученную нейросеть с вычисленными весами синапсов  $\alpha_i$ . Считаем, что погрешности входных сигналов, внутренних сигналов сети и элементов отсутствуют. При векторе входных сигналов  $\{z^{in}\}$  получаем вектор выходных сигналов  $\{y^{out}\}$ . Вектор  $\{y^{out}\}$  и внутренние сигналы сети  $x_i$  будем считать точным вектором выходных сигналов и точными сигналами сети.

Рассмотрим теперь эту же сеть, но предположим, что все сигналы сети имеют некоторые погрешности. Пусть  $\{y'\}$  – вектор выходных сигналов, полученный при том же векторе входных сигналов  $\{z^{in}\}$ , но с погрешностями внутренних сигналов сети.

Предполагаем, что внутри каждого слоя погрешности сигналов  $\varepsilon_i$  являются независимыми случайными величинами. Это предположение позволяет налагать менее жесткие требования при вычислении погрешностей сигналов.

Пусть нам задана  $\delta$  – допустимая погрешность выходных сигналов сети. То есть вектор  $\{y'\}$  может отличаться от вектора  $\{y^{out}\}$  не более, чем на  $\delta$ . Будем считать  $\delta$  величиной среднеквадратического отклонения  $\sigma_{out}$  выходных сигналов сети  $\{y'\}$ .

Нам нужно выяснить, каким образом могут распределяться дисперсии сигналов при заданном  $\sigma_{out}$  и вычислить среднеквадратические отклонения  $\sigma_i = \sqrt{D_i}$  для всех сигналов сети такие, чтобы среднеквадратическое отклонение вектора выходных сигналов  $\{y'\}$  равнялось  $\sigma_{out}$ .

Зная среднеквадратическое отклонение выходных сигналов, можем вычислить дисперсию выходных сигналов  $D_{out} = \sigma_{out}^2$ , а затем, переходя от

элемента к элементу в обратном порядке, вычислим дисперсии  $D_i$  и среднеквадратические отклонения  $\sigma_i = \sqrt{D_i}$  для всех сигналов сети.

Типичным участком сети является стандартный нейрон. Из стандартных нейронов состоит любая нейронная сеть. Поэтому нам достаточно определить, как вычисляются среднеквадратические отклонения сигналов для элементов стандартного нейрона. Тогда мы будем иметь возможность вычислить среднеквадратические отклонения для любого участка сети.

Выясним, как вычисляются среднеквадратические отклонения для входных сигналов точки ветвления, нелинейного преобразователя и сумматора, если нам будут известны среднеквадратические отклонения выходных сигналов этих элементов.

Если дисперсии выходных сигналов точки ветвления  $D_1, D_2, \dots, D_k$  при обратном распространении не равны между собой, то в качестве дисперсии входного сигнала точки ветвления выбирается  $\min\{D_i\}_{i=1}^k$ .

Пусть  $\sigma_1$  – среднеквадратическое отклонение погрешности выходного сигнала нелинейного преобразователя. Пусть случайная величина  $\varepsilon$  (погрешность входного сигнала нелинейного преобразователя) имеет некоторую плотность распределения  $\rho_\varepsilon$ . Считаем, что математическое ожидание погрешности входного сигнала  $M_\varepsilon = \int_{-\infty}^{\infty} \varepsilon \cdot \rho_\varepsilon d\varepsilon = 0$  и дисперсия

$$D_\varepsilon = \int_{-\infty}^{\infty} (\varepsilon - M_\varepsilon)^2 \rho_\varepsilon d\varepsilon = \sigma^2.$$

Пусть нелинейный преобразователь имеет функцию активации  $\varphi$  и точный входной сигнал  $A$ . Рассмотрим линейное приближение функции активации  $\varphi$  в точке  $A$ . Линейное приближение имеет вид:  $\varphi(A \pm \varepsilon) \approx \varphi(A) \pm \varphi'(A) \cdot \varepsilon$ . Найдем математическое ожидание и дисперсию величины  $\varphi(A + \varepsilon)$ .

$$M_{\varphi(A+\varepsilon)} \approx \int_{-\infty}^{\infty} (\varphi(A) + \varphi'(A) \cdot \varepsilon) \cdot \rho_{\varepsilon} d\varepsilon = \varphi(A).$$

$$D_{\varphi(A+\varepsilon)} \approx \int_{-\infty}^{\infty} (\varphi(A) + \varphi'(A) \cdot \varepsilon - \varphi(A))^2 \rho_{\varepsilon} d\varepsilon = \varphi'(A)^2 \sigma^2.$$

С другой стороны, нам известно, что дисперсия выходного сигнала нелинейного преобразователя равна  $\sigma_1^2$ . Отсюда получаем

$$\sigma_1^2 = \varphi'(A)^2 \cdot \sigma^2 \Rightarrow \sigma = \frac{\sigma_1}{|\varphi'(A)|}.$$

Таким образом, мы вычислили среднеквадратическое отклонение входного сигнала нелинейного преобразователя для любого распределения погрешности входного сигнала  $\varepsilon$ .

Мы получили среднеквадратическое отклонение входного сигнала нелинейного преобразователя  $\sigma$ , которое одновременно является среднеквадратическим отклонением выходного сигнала сумматора с погрешностями входных сигналов  $\sum_{i=1}^n \alpha_i \cdot (x_i + \varepsilon_i)$ . Погрешность выходного

сигнала сумматора равняется  $\sum_{i=1}^n \alpha_i \cdot (x_i + \varepsilon_i) - \sum_{i=1}^n \alpha_i \cdot x_i = \sum_{i=1}^n \alpha_i \cdot \varepsilon_i$ , где

$\sum_{i=1}^n \alpha_i \cdot x_i$  – точный выходной сигнал сумматора.

Вычислим среднеквадратические отклонения  $\sigma_i$  входных сигналов сумматора. Рассмотрим для этого дисперсию погрешности выходного сигнала

сумматора  $D(\sum_{i=1}^n \alpha_i \cdot \varepsilon_i) = \sigma^2$ .

Предположим дополнительно, что  $\sigma_i$  равны между собой.

$$\sigma^2 = D(\sum_{i=1}^n \alpha_i \cdot \varepsilon_i) = \sum_{i=1}^n D(\alpha_i \cdot \varepsilon_i) = \sum_{i=1}^n \alpha_i^2 \cdot D\varepsilon_i =$$

$$= \sum_{i=1}^n \alpha_i^2 \cdot \sigma_i^2 \Rightarrow \sigma_i = \frac{\sigma}{\sqrt{\sum_{i=1}^n \alpha_i^2}}.$$

Получили формулу для равномерного распределения среднеквадратических отклонений  $\sigma_i$  по входам сумматора. Если в качестве погрешности каждого входа рассматривать не  $\varepsilon_i$ , а  $\alpha_i \cdot \varepsilon_i$ , то получим формулу для пропорционального распределения среднеквадратических отклонений  $\sigma_i$  по входам сумматора.

$$\sigma^2 = D\left(\sum_{i=1}^n \alpha_i \cdot \varepsilon_i\right) = \sum_{i=1}^n D(\alpha_i \cdot \varepsilon_i) = \sum_{i=1}^n \sigma_i^2 \Rightarrow \sigma_i = \frac{\sigma}{\sqrt{n}}.$$