Воронов Евгений Алексеевич
магистрант
Новокузнецкий институт (филиал)
ФГБОУ ВО «Кемеровский государственный университет»
г. Новокузнецк, Кемеровская область

МЕТОД РЕГУЛЯРИЗАЦИИ В ЗАДАЧЕ ОБУЧЕНИЯ РАДИАЛЬНЫХ НЕЙРОННЫХ СЕТЕЙ

Аннотация*: в статье рассматриваются: теория регуляризации, метод негладкой регуляризации, сглаживающие функционалы для подавления избыточных переменных и принципы их построения.*

Ключевые слова*: нейронные сети, теория регуляризации, метод негладкой регуляризации, сглаживающие функционалы для подавления избыточных переменных, принципы построения сглаживающих функционалов, методы обучения нейронных сетей.*

Теория регуляризации

В 1963 году Тихонов предложил новый метод, получивший название регуляризации и предназначенный для решения задачи обучения нейронных сетей. В контексте задачи восстановления гиперповерхности главная идея регуляризации заключается в стабилизации решения с помощью некоторой вспомогательной неотрицательной функции, которая несет в себе априорную информацию о решении. Наиболее общей формой априорной информации является предположение о гладкости функции искомого отображения (т. е. решения задачи восстановления) в том смысле, что одинаковый входной сигнал соответствует одинаковому выходному.

Для примера возьмем множество пар данных «вход-выход» (т. е. пример обучения), доступных для аппроксимации и описываемых следующим образом.

Входной сигнал: $x_i \in R^{m_0}$, i= 1,2,..., N.

Желаемый отклик: $d_i \in R^1$,i=1,2,..., N.

Обозначим функцию аппроксимации как F(x), где (для упрощения выкладок) в списке аргументов опущен вектор весов w. Теория регуляризации Тихонова в своем изначальном виде использует два слагаемых.

Слагаемое стандартной ошибки. Первое слагаемое, обозначаемое E_S, описывает стандартную ошибку (расстояние между желаемым откликом d_i и фактическим выходным сигналом y_i для примера обучения i= 1, 2,..., N). В частности, можно определить

$$E_S(F) = \frac{1}{2}\sum_{i=1}^{N}(d_i - y_i)^2 = \frac{1}{2}\sum_{i=1}^{N}(d_i - F(x_i))^2 \quad (1)$$

1. Слагаемое регуляризации. Это второе слагаемое, обозначаемое E_C, зависит от «геометрических» свойств функции аппроксимации F(x).

В частности, можно записать:

$$E_C(F) = \frac{1}{2}\|DF\| \quad (2)$$

где D линейный дифференциальный оператор. Априорная информация о форме решения (т. е. о функции отображения F(x)), включенная в дифференциальный оператор D, обеспечивает его зависимость от конкретной задачи. Оператор D иногда еще называют стабилизатором, так как в задаче регуляризации он стабилизирует решение, делая его гладким и, таким образом, удовлетворяющим свойству непрерывности.

Символ ‖ • ‖ в выражении (2) обозначает норму в функциональном пространстве, к которому принадлежит DF(x). При обычных условиях используемое здесь функциональное пространство является пространством L_2, состоящим из всех действительных функций f(x), $x \in R^{m_0}$, для которых норма $\|f(x)\|^2$ является интегрируемой по Лебегу. Используемая здесь функция f(x) обозначает фактическую функцию, описывающую моделируемый физический процесс, отвечающий за генерацию пар примеров обучения $\{(x_i, d_i)\}_{i=1}^{N}$.

Величиной, которую требуется минимизировать, в теории регуляризации является

$$E(F) = E_S(F) + \alpha E_C(F) = \frac{1}{2}\sum_{i=1}^{N}(d_i - F(x_i))^2 + \frac{1}{2}\alpha \parallel DF \parallel^2 \qquad (3)$$

где α – положительное действительное число, называемое параметром регуляризации; E(F) – функционал Тихонова. Функционал отображает функции (определенные в соответствующем функциональном пространстве) на ось действительных чисел. Аргминимум функционала Тихонова E(F) (т.е. решение задачи регуляризации) обозначается $F_\alpha(x)$.

В некотором смысле параметр регуляризации α. можно рассматривать как индикатор достаточности данного набора данных для определения решения $F_\alpha(x)$. В частности, крайний случай, $\alpha \to 0$, означает, что задача является безусловной и имеет решение $F_\alpha(x)$, целиком зависящее от примеров. Другой крайний случай, $\alpha \to \infty$, предполагает, что самого априорного ограничения на гладкость, представленного дифференциальным оператором D, достаточно для определения решения $F_\alpha(x)$. Это может указывать также на недостоверное количество примеров. В практических приложениях параметр регуляризации α принимает некоторое среднее значение между этими двумя крайними случаями. Этим определяется влияние на решение $F_\alpha(x)$ как априорной информации, так и данных обучающей выборки. Таким образом, слагаемое регуляризации $E_C(F)$ представляет собой функцию штрафа за сложность модели, влияние которой на окончательное решение определяется параметром регуляризации α.

Метод негладкой регуляризации

Рассмотрим технологию метода негладкой регуляризации, которая заключается в отсеве неинформативных переменных линейной модели посредством последовательности чередующихся шагов ранжирования признаков средствами негладкой регуляризации по их значимости и удаления первых малозначимых признаков согласно некоторому критерию. Для подавления избыточных переменных предварительное обучение следует производить посредством минимизации по *w* квадратичной ошибки и негладкого сглаживающего функционала:

$$E_\Omega(\alpha, w, D) = \sum_{x,y\in D}(y - f(x,w))^2 + \alpha\Omega(w), \qquad (4)$$

$$\Omega(w) = \sum_{i \in Iw} \left| w_i \right|^{\gamma}, \; 0 < \gamma < 1, \tag{5}$$

где α – параметр регуляризации, Iw – множество номеров переменных массива w – по которым проводится регуляризация. Функционал $\Omega(w)$ предназначен для подавления избыточных переменных модели $f(x, w)$. Поэтому в решении будет присутствовать множество компонент близких к нулю, которые необходимо исключать, используя специальные алгоритмы.

Сглаживающие функционалы для подавления избыточных переменных и принципы их построения

Линии уровня первого слагаемого функционала (4) (линии уровня эллипсоидального вида) и функционала (5) (линии уровня звездоподобного вида) изображены на рисунке 1.

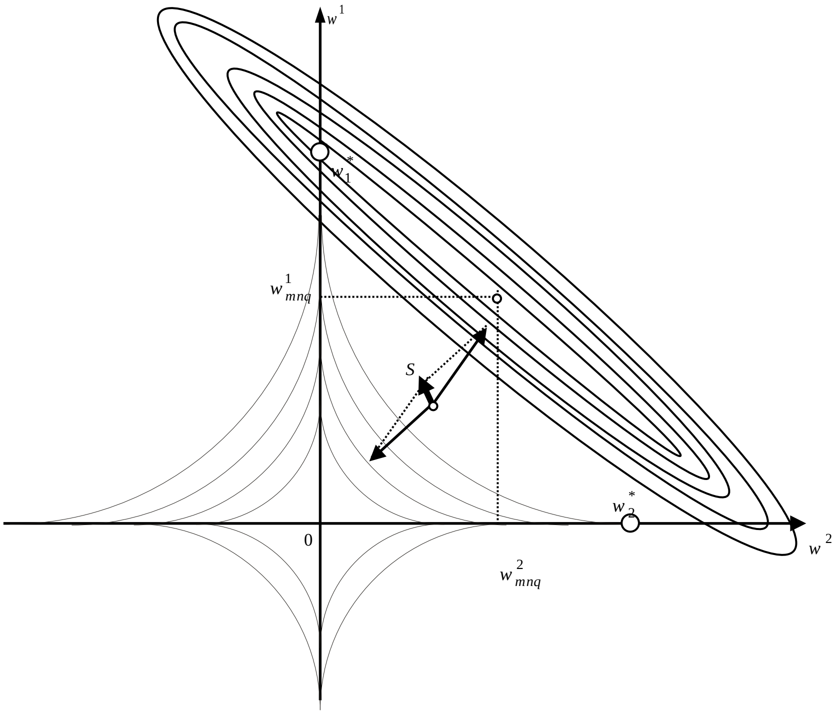

Рис. 1. Линии уровня функционалов

На рис. 1 справа сверху изображены линии уровня первого слагаемого функционала (4) а слева снизу – сглаживающего функционала (5). Точка w_{mnq} – точка минимума первого слагаемого (4). Функционал (5) осуществляет дополнительное притяжение отдельных компонент точки минимума к нулю. За счет того, что $0 < \gamma < 1$, результирующая функция (4) может иметь несколько локальных минимумов (на рис. 1 – это точки w_1^* и w_2^*, имеющие нулевые координаты). Как показывает вычислительный эксперимент, вследствие сложности оптимизационной задачи минимизации функции (4), ее локальные минимумы не всегда имеют нулевые компоненты, но всегда имеют компоненты близкие к нулю, что позволяет установить ранговое упорядочение компонент решения по их величине и отбросить часть признаков с малыми компонентами решения.