$$|a + b + \bar{a}| + c_p/c\sqrt{0.5\bar{a}} + (|ab| + b^2)\delta_0 + c_p/bc\sqrt{\delta_0} < \bar{a},$$

$$\text{где } \delta_0 = \max\{\log h_0, (1 - h_0^{-1})r\}$$

1. Н. В. Азбелев, П. М. Симонов. Устойчивость решений уравнений с обыкновенными производными. Пермь: Изд-во Перм. ун-та, 2001. 230 с.
2. Е. А. Барбашин. Введение в теорию устойчивости. М.: Наука, 1967. 224 с.
3. Р. Ш. Липцер, А. Н. Ширяев. Теория мартингалов. М.: Наука, 1986. 512 с.
4. Х. Л. Массера, Х. Х. Шефер. Линейные дифференциальные уравнения и функциональные пространства. М.: Мир. 1970. 456 с.
5. Е. Ф. Царьков. Случайные возмущения дифференциально-функциональных уравнений. Рига: Зинатне 1989. 421 с.
6. J. A. D. Appleby and D. W. Reynolds. Non-exponential stability of scalar stochastic Volterra equations // Statist. Probab. Lett. 2003. V. 62, no. 4. P. 335-343.
7. R. I. Kadiev and A. V. Ponosov. Stability of stochastic functional differential equations and the W-transform // Electron. J. Diff. Eqns. 2004. V. 2004, no. 92. P. 1-36.
8. V. B. Kolmanovskii and V. R. Nosov. Stability of Functional Differential Equations. NewYork: AcademicPress. 1986. 217 p.
9. X. Mao. Stochastic Differential Equations and Applications. Chichester: Horwood Publishing ltd. 1997. 360 p.

Самойленко Н.С.

Алгоритм обучения нейронных сетей на малых выборках

Кемеровский государственный университет
(Россия, Кемерово)

doi:10.18411/spc-12-06-2018-02

idsp: 000001:spc-12-06-2018-02

Аннотация

Традиционно возникают проблемы, связанные с избыточностью описания в случае большого числа нейронов и недостаточными возможностями обеспечения обобщающих свойств сети при малом их числе. В работе рассматриваются алгоритмы обучения двухслойных нейронных сетей при условии малого числа выборочных данных.

Ключевые слова: нейрон, нейронная сеть, регуляризация, обучение, метод минимизации.

Искусственная нейронная сеть (ИНС) [1–4] – один из способов реализации искусственного интеллекта. В различных приложениях аппарата ИНС [1] при небольшом числе данных и их неравномерном распределении в признаковом пространстве в задачах обучения [1] возникают проблемы, связанные с избыточностью описания в случае большого числа нейронов и недостаточными возможностями для обеспечения хороших обобщающих свойств [1] при малом их числе. За счет выбора хорошего начального приближения и разумного набора компонент нейронов удается и на малых данных получать качественные приближения данных ИНС с небольшим числом слоев. Представляется актуальной разработка алгоритмов автоматического построения начального приближения ИНС и методов выбора информативных компонент нейронов.

В работе рассматривается алгоритм обучения двухслойных ИНС сигмоидального типа. Для выбора начального приближения ИНС используется идея окаймления поверхности, задаваемой данными, поверхностями избыточного числа нейронов, распределенных некоторым методом кластеризации по области данных. Для устранения избыточного описания ИНС используют различные способы регуляризации [1, 5, 6]. В силу не гладкости используемых регуляризаторов в качестве алгоритма

обучения используется субградиентный метод минимизации с растяжением – сжатием пространства [10], принадлежащий классу релаксационных субградиентных методов.

Задача аппроксимации двухслойной ИНС сигмоидального типа. ИНС – мощный инструмент аппроксимации. Рассмотрим задачу аппроксимации [1]:

$$w^* = \operatorname{argmin}_w E(\alpha, w, D), \quad (1)$$

$$E(\alpha, w, D) = \sum_{x, y \in D} (y - f(x, w))^2 + \sum_{i=1}^k \alpha_i R_i(w),$$

где $D = \{(x^i, y_i) | x^i \in R^P, y_i \in R^1\}$, $i = 1, \dots, N$ – данные наблюдения, $R_i(w)$ – различные виды регуляризаторов, α_i – параметры регуляризации, $f(x, w)$ – аппроксимирующая функция, $x \in R^P$ – вектор данных, $w \in R^n$ – вектор настраиваемых параметров, p и n – их размерности. В качестве регуляризаторов используются:

$$R2(w) = \sum_{i=1}^n w_i^2, \quad R1(w) = \sum_{i=1}^n |w_i|,$$

$$R\gamma(w) = \sum_{i=1}^n (|w_i| + \varepsilon)^\gamma \quad (\varepsilon = 10^{-6}, \gamma = 0.7).$$

Использование $R2$ приводит к подавлению в большей мере больших компонент вектора w , $R1$ – больших и малых, а $R\gamma$ – преимущественно малых. Подобное свойство $R\gamma$ позволяет сводить к нулю слабые компоненты, несущественные для описания данных, и не влиять на информативные компоненты. Регуляризатор $R2$ эффективен при подавлении помех.

В задаче аппроксимации сетью прямого распространения требуется по данным D обучить двухслойную сигмоидальную нейронную сеть [2, 3] следующего вида (оценить ее неизвестные параметры w)

$$f(x, w) = w_0^{(2)} + \sum_{i=1}^m w_i^{(2)} \varphi(s_i), \quad (2)$$

где $s_i = w_{i0}^{(1)} + \sum_{j=1}^p x_j w_{ij}^{(1)}$, $i = 1, 2, \dots, m$, $\varphi(s) = s / (1 + |s|)$, x_j – компоненты

вектора $x \in R^P$, $w = ((w_i^{(2)}, i = 0, \dots, m), (w_{ij}^{(1)}, j = 0, \dots, p, i = 1, \dots, m))$ – набор неизвестных параметров, которые необходимо оценить методом наименьших квадратов (1), $\varphi(s)$ – функция активации нейрона, m – число нейронов.

Для решения оптимизационных задач (1) используем релаксационный субградиентный метод с растяжением – сжатием пространства (SMDM) [7]. На гладких функция метод является эффективным при минимизации негладких функций.

Алгоритм нахождения начального приближения. Рабочие области нейронов $\varphi(s)$ расположены в некоторой окрестности значений $s=0$, а в других областях близки к своим асимптотам. Стандартная практика заключается в выборе начального приближения случайным образом [2, 3]. При этом зачастую оказывается, что рабочие области нейронов охватывают только часть области аппроксимации, либо выходят за ее пределы. Расположение нейронов в точках с высокой концентрацией данных также не обеспечивает сохранения этого положения, поэтому требуется дополнительная привязка рабочих областей нейронов посредством обучения нейросети при фиксированных центрах.

В предлагаемом алгоритме находится приближение ИНС при фиксированном положении рабочих областей нейронов с помощью заданных центров $c_i \in R^P$, $i = 1, 2, \dots, m$, в области аппроксимации $x \in R^P$, определяемой данными. В

этом случае в нейронах исключен свободный член, а в (2) будет использоваться выражение:

$$s_i = \sum_{j=1}^P (x_j - c_{ij}) w_{ij}^{(1)}, \quad i=1,2,\dots,m. \quad (3)$$

где компоненты вектора w для нейронов не содержат свободных членов.

Центры c_i можно найти некоторым алгоритмом кластеризации данных $x^i \in R^P, i=1,\dots,N$, что обеспечит расположение нейронов в областях с высокой плотностью данных. Используем максиминный алгоритм, в котором в качестве первых двух центров выбираются две максимально удаленные друг от друга точки данных. Каждый новый центр получается выбором точки данных x^i , расстояние от которой до ближайшего известного центра максимально. После решения задачи (1) при фиксированных центрах необходимо вернуться к исходному описанию сети в виде (2) посредством образования свободного члена нейрона, а остальные параметры оставить без изменений:

$$w_{i0}^{(1)} = -\sum_{j=1}^P c_{ij} w_{ij}^{(1)}, \quad i=1,2,\dots,m.$$

Такой алгоритм отыскания начального приближения ИНС гарантирует, что область данных будет покрыта рабочими областями нейронов.

Заключение

В работе предложен алгоритм построения ИНС в условиях небольших по размеру обучающих выборках. Разработан алгоритм нахождения начального приближения ИНС. Высокая скорость сходимости алгоритма минимизации дает возможность эффективно решать задачи обучения ИНС с негладкой регуляризацией. Эффективность алгоритма обучения подтверждается на практических примерах.

1. Воронцов К.В. Курс лекций «Математические методы обучения по прецедентам» // <http://www.ccas.ru/voron/>, <http://www.machinelearning.ru/>.
2. Haykin S. Neural Networks: A Comprehensive Foundation. New York: MacMillan, 1994.
3. Osowski S. Neural Networks for Information Processing. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2000.
4. Горбань А. Н., Обучение нейронных сетей. М.: изд. СССР-США СП «Параграф», 1990. 160 с.
5. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data Mining, In-ference, and Prediction. Springer, 2009.
6. Крутиков В.Н., Арышев Д.В. Исследование субградиентных методов обучения нейронных сетей // Вестник КемГУ. Кемерово. 2004. №3(3). С. 119-123.
7. Крутиков, В.Н., Горская Т.А. Семейство релаксационных субградиентных методов с двухранговой коррекцией матриц метрики // Экономика и мат. методы. 2009. Т. 45. Вып. 4. С. 37-80.