Санкт-Петербургский государственный университет
Прикладная математика и информатика
Управление и обработка информации в кибернетических и робототехнических системах

Дорофеев Антон Сергеевич

Алгоритм скоростного градиента в задаче обучения нейросетей

Выпускная квалификационная работа

Научный руководитель:
к. ф.-м. н. М. С. Ананьевский

Рецензент:
д. т. н. И. Б. Фуртат

Санкт-Петербург
2018

Saint Petersburg State University
Applied Mathematics and Computer Science
Control and data processing in cybernetic and robotic systems

Dorofeev Anton

Speed-gradient algorithm for training artificial neural networks

Graduation Project

Scientific Supervisor:
Ph. D. M. S. Ananyevskiy

Reviewer:
Ph. D. I. B. Furtat

Saint Petersburg
2018

Оглавление

Введение

В данной работе рассматривается задача классификации динамических процессов по мгновенным наблюдениям. В качестве метода решения используется построение однослойного персептрона.

Если удается обучить персептрон, то, учитывая динамику процессов, через некоторое время он может начать делать неправильные выводы о процессах, на которых мы его обучили. Чтобы скорректировать вывод персептрона, будем корректировать веса, превратив их в функцию времени. Тогда персептрон будет меняться вместе с процессами и таким образом корректировать свой вывод с течением времени.

Глава 1

Постановка задачи

Имеется N процессов $x_1(t), \ldots, x_N(t)$, каждый из которых принадлежит либо классу C_A, либо классу C_B. Информация о принадлежности к классу неизвестна. Для каждого процесса наблюдается M величин:

$$y_1\big(x_k(t)\big), \ldots, y_M\big(x_k(t)\big), \quad k = 1, \ldots, K.$$

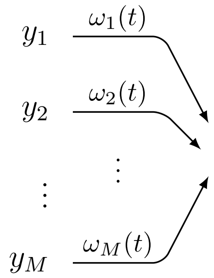
Требуется построить классификатор, который по вектору наблюдений $\Big[y_1\big(x(t_0)\big), \ldots, y_M\big(x(t_0)\big)\Big]^{\mathrm{T}}$, будет относить процесс $x(t)$ к его классу.

Глава 2

Решение

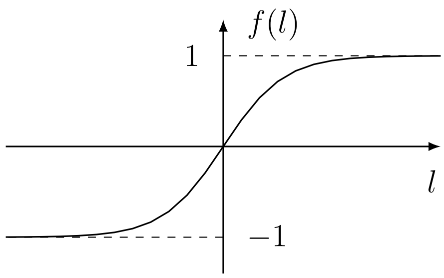
2.1. Метод решения

Построение однослойного персептрона с весами, которые есть функции времени: $\omega(t) = \big[\omega_1(t), \dots, \omega_M(t)\big]$.



$$z = f\big(y_1\omega_1(t) + y_2\omega_2(t) + \ldots + y_M\omega_M(t)\big) = \begin{cases} \geq 0 \;\Rightarrow\; C_A \\ < 0 \;\Rightarrow\; C_B \end{cases},$$

где $f(l) = \dfrac{2}{1+e^{-l}} - 1$ — функция активации нейрона(сигмоида).



2.2. Функция ошибки

$x_1(t), \dots, x_S(t)$ — обучающая выборка.

$$Q(t) = Q\Big(z\big(x_1(t)\big), \dots, z\big(x_S(t)\big), t\Big) = \frac{1}{2}\sum_{k=1}^{S}\Big[z\big(x_k(t)\big) - z^*\big(x(\cdot)\big)\Big]^2, \tag{2.1}$$

где $z\big(x(t)\big) = z\Big(y\big(x(t)\big), \omega(t)\Big) = f\Big(y_1\big(x(t)\big)\omega_1(t) + \ldots + y_M\big(x(t)\big)\omega_M(t)\Big)$,

$$z^*\big(x(\cdot)\big) = \begin{cases} 1, \; x(\cdot) \in C_A \\ -1, \; x(\cdot) \in C_B \end{cases}.$$

$\omega(t)$ ищется в аффинном виде: $\omega(t) = \omega^{BP} + \omega^{SG}t$, где ω^{BP} — весовая функция, найденная по методу Backpropogation, а ω^{SG} — весовой коэффициент, найденный по методу скоростного градиента исходя из ограничений $\dot{Q}(0) = \dot{Q}_{\max}$, где $\dot{Q}_{\max}$ — наперёд заданная константа. В этом случае сделаем предположение, что с такой временной поправкой целевая функция $Q(t)$ будет дольше находиться в желаемой зоне, чем если бы этой поправки не было.

2.2.1. Весовая функция

$$\omega^{BP} = \left[\omega_1^{BP}, \ldots, \omega_M^{BP}\right], \quad L(x) = \sum_{j=1}^{M} y_j(x)\omega_j^{BP}. \tag{2.2}$$

Поиск весовых функций производится в направлении антиградиента с шагом:

$$\Delta\omega_j^{BP} = -\lambda \frac{\partial Q(t)}{\partial \omega_j^{BP}}, \quad 0 < \lambda < 1,\ j = 1, \ldots, M. \tag{2.3}$$

Воспользуемся правилом дифференцирования сложных функций. Функция ошибки зависит в первую очередь от выходных сигналов персептрона:

$$\frac{\partial Q(t)}{\partial \omega_j^{BP}} = \sum_{k=1}^{S} \frac{\partial Q(t)}{\partial z\big(x_k(t)\big)} \cdot \frac{\partial z\big(x_k(t)\big)}{\partial \omega_j^{BP}} = \sum_{k=1}^{S} \Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big) \frac{\partial z\big(x_k(t)\big)}{\partial \omega_j^{BP}}. \tag{2.4}$$

В свою очередь, ω_j^{BP} влияет на выход сети только как часть суммы $L(x)$, следовательно

$$\frac{\partial z\big(x_k(t)\big)}{\partial \omega_j^{BP}} = \frac{\partial z\Big(y\big(x_k(t)\big), \omega(t)\Big)}{\partial \omega_j^{BP}} = \left(\frac{\partial f(s)}{\partial s}\right)\bigg|_{l=L(x_k)} \frac{\partial L(x_k)}{\partial \omega_j^{BP}} =$$
$$= f\big(L(x_k)\big)\Big(1 - f\big(L(x_k)\big)\Big) y_j(x_k) = y_j(x_k) \frac{2e^{-L(x_k)}}{(1 + e^{-L(x_k)})^2}. \tag{2.5}$$

Подставим полученный результат и найдем $\Delta\omega_j^{BP}$:

$$\frac{\partial Q(t)}{\partial \omega_j^{BP}} = \sum_{k=1}^{S} \Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big) y_j(x_k) \frac{2e^{-L(x_k)}}{(1 + e^{-L(x_k)})^2}, \tag{2.6}$$

$$\Delta\omega_j^{BP} = -\lambda \sum_{k=1}^{S} \Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big) y_j(x_k) \frac{2e^{-L(x_k)}}{(1 + e^{-L(x_k)})^2}. \tag{2.7}$$
$$0 < \lambda < 1, \quad j = 1, \ldots, M$$

Организуем итерационную процедуру поиска весовых функций с шагом $\Delta\omega_j^{BP}$ и найдем ω^{BP}, минимизирующий $Q(t)$. Считаем, что поиск осуществляется мгновенно и $Q(0) = 0$.

2.2.2. Весовой коэффициент

Напишем формулу производной функции ошибки:

$$\dot{Q}(t) = \frac{dQ(t)}{dt} = \sum_{k=1}^{S} \frac{\partial Q(t)}{\partial z} \cdot \frac{dz\big(x_k(t)\big)}{dt} = \sum_{k=1}^{S} \Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big) \cdot$$
$$\cdot \sum_{j=1}^{M} \left(\frac{\partial z\Big(y\big(x_k(t)\big), \omega(t)\Big)}{\partial y_j} \cdot \frac{dy_j\big(x_k(t)\big)}{dt} + \frac{\partial z\Big(y\big(x_k(t)\big), \omega(t)\Big)}{\partial \omega_j} \cdot \frac{d\omega_j(t)}{dt} \right) =$$
$$= \sum_{k=1}^{S} \Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k(t))}}{(1+e^{-L(x_k(t))})^2} \sum_{j=1}^{M} \big(\omega_j(t)\dot{y}_j\big(x_k(t)\big) + y_j\big(x_k(t)\big)\dot{\omega}_j(t)\big). \quad (2.8)$$

$$\omega(t) = \omega^{BP} + \omega^{SG} t \quad \Longrightarrow \quad \begin{cases} \dot{\omega}(t) = \omega^{SG} & \Longrightarrow \quad \dot{\omega}_j(t) = \omega_j^{SG} \\ \omega(0) = \omega^{BP} & \Longrightarrow \quad Q(0) = 0 \end{cases}.$$

$$\dot{Q}(t) = \sum_{k=1}^{S} \Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k(t))}}{(1+e^{-L(x_k(t))})^2} \sum_{j=1}^{M} \big(\omega_j(t)\dot{y}_j\big(x_k(t)\big) + y_j\big(x_k(t)\big)\omega_j^{SG}\big). \quad (2.9)$$

Необходимо найти весовой коэффициент $\omega^{SG} = \big[\omega_1^{SG}, \ldots, \omega_M^{SG}\big]$, решающий задачу минимизации функционала

$$J\big(\dot{Q}(0)\big) = \frac{1}{2}\left(\dot{Q}\,(0) - \dot{Q}_{\max}\right)^2, \text{ где}$$
$$\dot{Q}(0) = \sum_{k=1}^{S} \Big(z(x_k) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k)}}{(1+e^{-L(x_k)})^2} \sum_{j=1}^{M} \big(\omega_j^{BP}\dot{y}_j(x_k) + y_j(x_k)\omega_j^{SG}\big), \ x_k(0) = x_k. \quad (2.10)$$

Поиск осуществляется в направлении антиградиента с шагом:

$$\Delta\omega_i^{SG} = -\gamma \frac{\partial J\big(\dot{Q}(0)\big)}{\partial \omega_i^{SG}} = -\gamma \frac{\partial J\big(\dot{Q}\,(0)\,\big)}{\partial\big(\dot{Q}\,(0)\,\big)} \cdot \frac{\partial \dot{Q}\,(0)}{\partial \omega_i^{SG}} = -\gamma\big(\dot{Q}\,(0) - \dot{Q}_{\max}\big)\frac{\partial \dot{Q}\,(0)}{\partial \omega_i^{SG}}. \quad (2.11)$$
$$0 < \gamma < 1, \quad i = 1, \ldots, M$$

$$\frac{\partial \dot{Q}\,(0)}{\partial \omega_i^{SG}} = \sum_{k=1}^{S} \Big[\frac{\partial z\big(x_k\big)}{\partial \omega_i^{SG}} \cdot \frac{2e^{-L(x_k)}}{(1+e^{-L(x_k)})^2} \sum_{j=1}^{M} \big(\omega_j^{BP}\dot{y}_j(x_k) + y_j(x_k)\omega_j^{SG}\big) +$$
$$+ \Big(z(x_k) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k)}}{(1+e^{-L(x_k)})^2} y_i(x_k) \Big] \quad (2.12)$$

$$\frac{\partial z\big(x_k\big)}{\partial \omega_i^{SG}} = \left(\frac{\partial f(l)}{\partial l}\right)\Bigg|_{l=L(x_k)} y_i(x_k) \cdot 0 = 0, \quad \Longrightarrow$$
$$\frac{\partial \dot{Q}\,(0)}{\partial \omega_i^{SG}} = \sum_{k=1}^{S} \Big(z(x_k) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k)}}{(1+e^{-L(x_k)})^2} y_i(x_k). \quad (2.13)$$

$$\Delta\omega_i^{SG} = -\gamma\big(\dot{Q}\,(0) - \dot{Q}_{\max}\big) \sum_{k=1}^{S} \Big(z(x_k) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k)}}{(1+e^{-L(x_k)})^2} y_i(x_k) =$$

$$= -\gamma \left[\sum_{p=1}^{S} \Big(z(x_p) - z^*\big(x_p(\cdot)\big)\Big) \frac{2e^{-L(x_p)}}{(1+e^{-L(x_p)})^2} \sum_{j=1}^{M} \big(\omega_j^{BP} \dot{y}_j(x_p) + y_j(x_p)\omega_j^{SG}\big) - \dot{Q}_{\max} \right] \cdot$$

$$\cdot \sum_{k=1}^{S} \Big(z(x_k) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k)}}{(1+e^{-L(x_k)})^2} y_i(x_k), \quad i = 1, \ldots, M. \quad (2.14)$$

Как и для поиска ω^{BP} организуем итерационную процедуру поиска весовых коэффициентов с шагом $\Delta\omega_j^{SG}$ и найдем ω^{SG}, минимизирующий $J\big(\dot{Q}(0)\big)$. Считаем, что поиск весовых коэффициентов осуществляется мгновенно.

2.2.3. Оценка функции ошибки

Разложим функционал ошибки в ряд Тейлора с остаточным членом в форме Лагранжа:

$$Q(\tau) = Q(0) + \tau\dot{Q}(0) + \frac{\tau^2}{2}\ddot{Q}(\Theta\tau), \quad 0 \le \Theta \le 1, \quad (2.15)$$

$$Q(\tau) = \tau\dot{Q}(0) + \frac{\tau^2}{2}\ddot{Q}(\Theta\tau), \quad 0 \le \Theta \le 1. \quad (2.16)$$

Воспользуемся формулой для $\dot{Q}(t)$ и напишем как будет выглядеть $\ddot{Q}(t)$:

$$\ddot{Q}(t) = \sum_{k=1}^{S} \frac{\partial \dot{Q}(t)}{\partial z} \cdot \frac{dz\big(x_k(t)\big)}{dt} + \sum_{k=1}^{S}\sum_{j=1}^{M} \Big(\frac{\partial \dot{Q}(t)}{\partial y_j} \cdot \frac{dy_j\big(x_k(t)\big)}{dt} + \frac{\partial \dot{Q}(t)}{\partial \omega_j} \cdot \frac{d\omega_j(t)}{dt} \Big), \quad (2.17)$$

$$\frac{2e^{-L(x_k)}}{(1+e^{-L(x_k)})^2} = z\big(x_k(t)\big)\Big(1 - z\big(x_k(t)\big)\Big), \quad (2.18)$$

$$\left(\frac{\partial \dot{Q}(t)}{\partial z}\right)\Bigg|_{z=z(x_k(t))} = \Big[-3z\big(x_k(t)\big)^2 + 2z\big(x_k(t)\big)\Big(z^*\big(x_k(\cdot)\big) + 1\Big) - z^*\big(x_k(\cdot)\big)\Big] \cdot$$

$$\cdot \sum_{j=1}^{M} \left(\omega_j(t)\dot{y}_j\big(x_k(t)\big) + y_j\big(x_k(t)\big)\omega_j^{SG}\right), \quad (2.19)$$

$$\frac{dz\big(x_k(t)\big)}{dt} = z\big(x_k(t)\big)\Big(1 - z\big(x_k(t)\big)\Big) \sum_{j=1}^{M} \left(\omega_j(t)\dot{y}_j\big(x_k(t)\big) + y_j\big(x_k(t)\big)\omega_j^{SG}\right), \quad (2.20)$$

$$\sum_{k=1}^{S}\sum_{j=1}^{M} \Big(\frac{\partial \dot{Q}(t)}{\partial y_j} \cdot \frac{dy_j\big(x_k(t)\big)}{dt} + \frac{\partial \dot{Q}(t)}{\partial \omega_j} \cdot \frac{d\omega_j(t)}{dt} \Big) =$$

$$= 2\sum_{k=1}^{S} \Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big) \frac{2e^{-L(x_k(t))}}{(1+e^{-L(x_k(t))})^2} \sum_{j=1}^{M} \dot{y}_j\big(x_k(t)\big)\omega_j^{SG}. \quad (2.21)$$

Подставим полученные результаты и найдем $\ddot{Q}(t)$:

$$\ddot{Q}(t) = \sum_{k=1}^{S} \Big[3z\big(x_k(t)\big)^4 - \Big(2z^*\big(x_k(\cdot)\big)+5\Big)z\big(x_k(t)\big)^3 + \Big(z^*\big(x_k(\cdot)\big)+2\Big)z\big(x_k(t)\big)^2 - \\ - z^*\big(x_k(\cdot)\big)z\big(x_k(t)\big)\Big]\left(\sum_{j=1}^{M}\big(\omega_j(t)\dot{y}_j\big(x_k(t)\big) + y_j\big(x_k(t)\big)\omega_j^{SG}\big)\right)^2 + \\ + 2\sum_{k=1}^{S}\Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big)\frac{2e^{-L(x_k(t))}}{(1+e^{-L(x_k(t))})^2}\sum_{j=1}^{M}\dot{y}_j\big(x_k(t)\big)\omega_j^{SG}. \quad (2.22)$$

Заметим, что

$$\Big[3z\big(x_k(t)\big)^4 - \Big(2z^*\big(x_k(\cdot)\big)+5\Big)z\big(x_k(t)\big)^3 + \\ + \Big(z^*\big(x_k(\cdot)\big)+2\Big)z\big(x_k(t)\big)^2 - z^*\big(x_k(\cdot)\big)z\big(x_k(t)\big)\Big] \le 14, \quad (2.23)$$

$$\Big(z\big(x_k(t)\big) - z^*\big(x_k(\cdot)\big)\Big)\frac{2e^{-L(x_k(t))}}{(1+e^{-L(x_k(t))})^2} \le 4, \quad (2.24)$$

при $-1 < z\big(x_k(t)\big) < 1,\ z^*\big(x_k(\cdot)\big) = \{0,1\}$. Пусть $\omega_j^{SG} > 0,\ j = 1,\dots,M$. Тогда

$$\ddot{Q}(t) \le 14\sum_{k=1}^{S}\left(\sum_{j=1}^{M}\big(\omega_j(t)\dot{y}_j\big(x_k(t)\big) + y_j\big(x_k(t)\big)\omega_j^{SG}\big)\right)^2 + 8\sum_{k=1}^{S}\sum_{j=1}^{M}\dot{y}_j\big(x_k(t)\big)\omega_j^{SG}. \quad (2.25)$$

Подставим оценку $\ddot{Q}(t)$ в разложение в ряд Тейлора и получим оценку $Q(\tau)$.

$$Q(\tau) \le \tau\dot{Q}_{\max} + \tau^2\sum_{k=1}^{S}\left[7\left(\sum_{j=1}^{M}\big(\omega_j(\Theta\tau)\dot{y}_j\big(x_k(\Theta\tau)\big) + y_j\big(x_k(\Theta\tau)\big)\omega_j^{SG}\big)\right)^2 + \\ + 4\sum_{j=1}^{M}\dot{y}_j\big(x_k(\Theta\tau)\big)\omega_j^{SG}\right], \quad \forall\ \Theta \in [0;1]. \quad (2.26)$$

Заключение

Результатом данной работы является алгоритм построения однослойного персептрона, решающего задачу классификации динамических процессов по мгновенным наблюдениям. Найдена оценка функции ошибки. Из полученной оценки можно сделать предположение, что целевая функция $Q(t)$ с временной поправкой на веса $\omega_j^{SG}t$ будет дольше находиться в желаемой зоне, чем без этой поправки.

Список литературы

1. Мирошник И.В., Никифоров В.О., Фрадков А.Л., Нелинейное и адаптивное управление сложными динамическими системами. СПб.: Наука, 2000.
2. Гелиг А.Х., Матвеев А.С., Введение в математическую теорию обучаемых распознающих систем и нейронных сетей. СПб.: Издательство СПбГУ, 2014.
3. Rumelhart D.E., Hinton G.E., Williams R.J., Learning Internal Representations by Error Propagation. Cambridge, MA: MIT Press, 1986.