

Санкт-Петербургский государственный университет  
Прикладная математика и информатика  
Динамические системы, эволюционные уравнения, экстремальные задачи и  
математическая кибернетика

Дорофеев Антон Сергеевич

АЛГОРИТМ СКОРОСТНОГО ГРАДИЕНТА ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ  
ДИНАМИЧЕСКИХ ОБЪЕКТОВ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Выпускная квалификационная работа

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

Рецензент:  
д. т. н. И. Б. Фуртат

Санкт-Петербург

2020

Saint Petersburg State University  
Applied Mathematics and Computer Science  
Dynamic systems, evolution equations, extremal problems and mathematical cybernetics

Dorofeev Anton

SPEED-GRADIENT ALGORITHM FOR THE PROBLEM OF CLASSIFYING  
DYNAMIC OBJECTS USING NEURAL NETWORKS

Graduation Project

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

Reviewer:  
Ph. D. I. B. Furtat

Saint Petersburg  
2020

# Оглавление

<b>Введение</b> . . . . .	4
<b>Глава 1. Постановка задачи</b> . . . . .	5
<b>Глава 2. Решение</b> . . . . .	6
2.1. Метод решения . . . . .	6
2.2. Обучающая выборка . . . . .	6
2.3. Функция ошибки . . . . .	7
2.3.1. Весовая функция . . . . .	7
2.3.2. Весовой коэффициент . . . . .	8
2.4. Обучение классификатора . . . . .	9
2.5. Классификация . . . . .	9
<b>Глава 3. Моделирование</b> . . . . .	11
<b>Заключение</b> . . . . .	15
<b>Список литературы</b> . . . . .	16

## Введение

В работе рассматривается задача классификации роя динамических объектов по наблюдениям, сделанным в дискретные моменты времени, с использованием однослойного перцептрона [1]. Для классификации статических процессов с успехом применяются стандартные методы построения перцептронов и их обучения, например, метод обратного распространения ошибки [2]. Однако в случае динамических процессов такой подход не будет учитывать их динамику, и качество классификации будет снижаться со временем. В предложенном методе корректировка вывода перцептрона достигается путем введения весов в виде аффинных функций, определяемых методом градиентной обратной связи [3]. Благодаря введению динамически изменяющихся весов метод учитывает динамику процессов. Предполагается, что с такой временной поправкой целевая функция  $Q(t)$  будет дольше находиться в желаемой зоне, т.е. качество классификации будет снижаться медленнее.

## Глава 1

**Постановка задачи**

Имеется  $N$  динамических объектов  $X(t) = \{x_1(t), \dots, x_N(t)\}$ . Для каждого объекта наблюдается  $M$  величин:

$$y_1(X(t_0)), \dots, y_M(X(t_0)),$$

где  $y_k(X(t_0)) = \{y_k(x_1(t_0)), \dots, y_k(x_N(t_0))\}$ ,  $k = 1, \dots, M$ . В момент времени  $t$  наблюдается  $N$  величин:

$$y(X(t)) = \{y(\hat{x}_1(t)), \dots, y(\hat{x}_N(t))\},$$

где  $\{x_1(t), \dots, x_N(t)\} = \{\hat{x}_1(t), \dots, \hat{x}_N(t)\}$ .

Требуется определить взаимно-однозначное соответствие между наблюдениями  $y(X(t))$  и объектами  $X(t)$ .

## Глава 2

## Решение

## 2.1. Метод решения

Построение однослойного перцептрона с весами, представленными в виде функций времени:  $\omega(t) = [\omega_1(t), \dots, \omega_M(t)] \in \mathbb{R}^M$ .

$$\begin{array}{c}
 y_1 \xrightarrow{\omega_1(t)} \\
 y_2 \xrightarrow{\omega_2(t)} \\
 \vdots \\
 y_M \xrightarrow{\omega_M(t)}
 \end{array}
 \begin{array}{c}
 \searrow \\
 \searrow \\
 \nearrow
 \end{array}
 z = f(y_1\omega_1(t) + y_2\omega_2(t) + \dots + y_M\omega_M(t)) = \begin{cases} \geq 0 & \Rightarrow C_A \\ < 0 & \Rightarrow C_B \end{cases},$$

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

## 2.2. Обучающая выборка

Пусть  $Y(\cdot) = (y_1(\cdot), \dots, y_M(\cdot)) \in \mathbb{R}^M$  — последовательность наблюдений.  $Y(\cdot)$  принадлежит классу  $C_A$ , если  $y_k(\cdot)$  — наблюдение объекта  $x_j(t)$ ,  $\forall k \in \{1, \dots, M\}$ ,  $j \in \{1, \dots, N\}$ .  $Y(\cdot)$  принадлежит классу  $C_B$ , если  $\exists m \in \{1, \dots, M\} : y_m(\cdot)$  — наблюдение объекта  $x_i(t)$ ,  $i \in \{1, \dots, N\}$ ,  $y_k(\cdot)$  — наблюдение объекта  $x_j(t)$ ,  $\forall k \in \{1, \dots, M\} \setminus \{m\}$ ,  $j \in \{1, \dots, N\} \setminus \{i\}$ .

Элементы обучающей выборки класса  $C_A$  обозначим  $Y_{C_A}(t)$ :

$$Y_i(t) = (y_1(x_i(t)), \dots, y_M(x_i(t))), \quad i \in \{1, \dots, N\}.$$

Элементы обучающей выборки класса  $C_B$  обозначим  $Y_{C_B}(t)$ :

$$\begin{aligned}
 Y_{ij}(t) &= (y_1(x_i(t)), \dots, y_{M-1}(x_i(t)), y_M(x_j(t))), \\
 i, j &\in \{1, \dots, N\}, \quad i \neq j, \forall Y_{ij}, Y_{km} : i \neq k \Rightarrow j \neq m.
 \end{aligned}$$

$Y_{C_A}(t) \cup Y_{C_B}(t) = \{\mathbb{Y}_1(t), \dots, \mathbb{Y}_{2N}(t)\}$  — обучающая выборка,  $\mathbb{Y}_k(t) = (y_{k1}(t), \dots, y_{kM}(t))$ ,  $k = 1, \dots, 2N$ .

## 2.3. Функция ошибки

$$Q(t) = Q\left(z(\mathbb{Y}_1(t)), \dots, z(\mathbb{Y}_{2N}(t)), t\right) = \frac{1}{2} \sum_{k=1}^{2N} \left[ z(\mathbb{Y}_k(t)) - z^*(\mathbb{Y}_k(\cdot)) \right]^2, \quad (2.1)$$

где  $z(\mathbb{Y}_k(t)) = z(\mathbb{Y}_k(t), \omega(t)) = f(y_{k1}(t)\omega_1(t) + \dots + y_{kM}(t)\omega_M(t))$ ,

$$z^*(\mathbb{Y}_k(\cdot)) = \begin{cases} 1, & \mathbb{Y}_k(\cdot) \in C_A \\ -1, & \mathbb{Y}_k(\cdot) \in C_B \end{cases}.$$

$\omega(t)$  ищется в аффинном виде:  $\omega(t) = \omega^{BP} + \omega^{SG}t$ , где  $\omega^{BP}$  — весовая функция, найденная по методу обратного распространения ошибки [2], а  $\omega^{SG}$  — весовой коэффициент, найденный по алгоритму скоростного градиента [4], исходя из ограничений  $\dot{Q}(0) = \dot{Q}_{\max}$ , где  $\dot{Q}_{\max}$  — заранее заданная константа. Порог  $\dot{Q}_{\max}$  задается экспертом из общего представления о системе, его физический смысл — предельное значение скорости ухудшения классификатора, т.е. как быстро могут расти ошибки классификации с течением времени.

### 2.3.1. Весовая функция

$$\omega^{BP} = [\omega_1^{BP}, \dots, \omega_M^{BP}] \in \mathbb{R}^M, \quad L(\mathbb{Y}_k) = \sum_{j=1}^M y_{kj} \omega_j^{BP}. \quad (2.2)$$

Поиск весовых функций производится в направлении антиградиента с шагом:

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

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

$$\frac{\partial Q(t)}{\partial \omega_j^{BP}} = \sum_{k=1}^{2N} \frac{\partial Q(t)}{\partial z(\mathbb{Y}_k(t))} \cdot \frac{\partial z(\mathbb{Y}_k(t))}{\partial \omega_j^{BP}} = \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k(t)) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{\partial z(\mathbb{Y}_k(t))}{\partial \omega_j^{BP}}. \quad (2.4)$$

В свою очередь,  $\omega_j^{BP}$  влияет на выход сети только как часть суммы  $L(\mathbb{Y}_k)$ , следовательно

$$\begin{aligned} \frac{\partial z(\mathbb{Y}_k(t))}{\partial \omega_j^{BP}} &= \frac{\partial z(\mathbb{Y}_k(t), \omega(t))}{\partial \omega_j^{BP}} = \left( \frac{\partial f(l)}{\partial l} \right) \Big|_{l=L(\mathbb{Y}_k)} \frac{\partial L(\mathbb{Y}_k)}{\partial \omega_j^{BP}} = \\ &= f(L(\mathbb{Y}_k)) \left( 1 - f(L(\mathbb{Y}_k)) \right) y_{kj} = y_{kj} \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2}. \end{aligned} \quad (2.5)$$

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

$$\frac{\partial Q(t)}{\partial \omega_j^{BP}} = \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k(t)) - z^*(\mathbb{Y}_k(\cdot)) \right) y_{kj} \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2}, \quad (2.6)$$

$$\Delta\omega_j^{BP} = -\lambda \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k(t)) - z^*(\mathbb{Y}_k(\cdot)) \right) \mathbb{Y}_{kj} \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2}. \quad (2.7)$$

$$0 < \lambda < 1, \quad j = 1, \dots, M$$

Поиск весовых функций  $\omega^{BP}$ , минимизирующих  $Q(t)$ , осуществляется итерационно с шагом  $\Delta\omega_j^{BP}$ .

### 2.3.2. Весовой коэффициент

Напишем формулу производной функции ошибки:

$$\begin{aligned} \dot{Q}(t) &= \frac{dQ(t)}{dt} = \sum_{k=1}^{2N} \frac{\partial Q(t)}{\partial z} \cdot \frac{dz(\mathbb{Y}_k(t))}{dt} = \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k(t)) - z^*(\mathbb{Y}_k(\cdot)) \right) \cdot \\ &\quad \cdot \sum_{j=1}^M \left( \frac{\partial z(\mathbb{Y}_k(t), \omega(t))}{\partial \mathbb{Y}_{kj}} \cdot \frac{d\mathbb{Y}_{kj}(t)}{dt} + \frac{\partial z(\mathbb{Y}_k(t), \omega(t))}{\partial \omega_j} \cdot \frac{d\omega_j(t)}{dt} \right) = \\ &= \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k(t)) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_k(t))}}{(1 + e^{-L(\mathbb{Y}_k(t))})^2} \sum_{j=1}^M (\omega_j(t) \dot{\mathbb{Y}}_{kj}(t) + \mathbb{Y}_{kj}(t) \dot{\omega}_j(t)). \end{aligned} \quad (2.8)$$

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

$$\dot{Q}(t) = \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k(t)) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_k(t))}}{(1 + e^{-L(\mathbb{Y}_k(t))})^2} \sum_{j=1}^M (\omega_j(t) \dot{\mathbb{Y}}_{kj}(t) + \mathbb{Y}_{kj}(t) \dot{\omega}_j^{SG}).$$

Необходимо найти весовой коэффициент  $\omega^{SG} = [\omega_1^{SG}, \dots, \omega_M^{SG}] \in \mathbb{R}^M$ , решающий задачу минимизации функционала

$$\begin{aligned} J(\dot{Q}(0)) &= \frac{1}{2} \left( \dot{Q}(0) - \dot{Q}_{\max} \right)^2, \quad \text{где} \\ \dot{Q}(0) &= \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2} \sum_{j=1}^M (\omega_j^{BP} \dot{\mathbb{Y}}_{kj} + \mathbb{Y}_{kj} \omega_j^{SG}), \\ \mathbb{Y}_k(0) &= \mathbb{Y}_k, \quad \mathbb{Y}_{kj}(0) = \mathbb{Y}_{kj}, \quad \dot{\mathbb{Y}}_{kj}(0) = \dot{\mathbb{Y}}_{kj}. \end{aligned} \quad (2.9)$$

Поиск осуществляется в направлении антиградиента с шагом:

$$\Delta\omega_i^{SG} = -\gamma \frac{\partial J(\dot{Q}(0))}{\partial \omega_i^{SG}} = -\gamma \frac{\partial J(\dot{Q}(0))}{\partial (\dot{Q}(0))} \cdot \frac{\partial \dot{Q}(0)}{\partial \omega_i^{SG}} = -\gamma (\dot{Q}(0) - \dot{Q}_{\max}) \frac{\partial \dot{Q}(0)}{\partial \omega_i^{SG}}. \quad (2.10)$$

$$0 < \gamma < 1, \quad i = 1, \dots, M$$



$$\frac{\partial \dot{Q}(0)}{\partial \omega_i^{SG}} = \sum_{k=1}^{2N} \left[ \frac{\partial z(\mathbb{Y}_k)}{\partial \omega_i^{SG}} \cdot \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2} \sum_{j=1}^M (\omega_j^{BP} \dot{y}_{kj} + y_{kj} \omega_j^{SG}) + \left( z(\mathbb{Y}_k) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2} y_{ki} \right]. \quad (2.11)$$

$$\begin{aligned} \frac{\partial z(\mathbb{Y}_k)}{\partial \omega_i^{SG}} &= \left( \frac{\partial f(l)}{\partial l} \right) \Big|_{l=L(\mathbb{Y}_k)} y_{ki} \cdot 0 = 0, \implies \\ \frac{\partial \dot{Q}(0)}{\partial \omega_i^{SG}} &= \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2} y_{ki}. \end{aligned} \quad (2.12)$$

$$\begin{aligned} \Delta \omega_i^{SG} &= -\gamma (\dot{Q}(0) - \dot{Q}_{\max}) \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2} y_{ki} = \\ &= -\gamma \left[ \sum_{p=1}^{2N} \left( z(\mathbb{Y}_p) - z^*(\mathbb{Y}_p(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_p)}}{(1 + e^{-L(\mathbb{Y}_p)})^2} \sum_{j=1}^M (\omega_j^{BP} \dot{y}_{pj} + y_{pj} \omega_j^{SG}) - \dot{Q}_{\max} \right] \cdot \\ &\quad \cdot \sum_{k=1}^{2N} \left( z(\mathbb{Y}_k) - z^*(\mathbb{Y}_k(\cdot)) \right) \frac{2e^{-L(\mathbb{Y}_k)}}{(1 + e^{-L(\mathbb{Y}_k)})^2} y_{ki}, \quad i = 1, \dots, M. \end{aligned} \quad (2.13)$$

Поиск весовых коэффициентов  $\omega^{SG}$ , минимизирующих  $J(\dot{Q}(0))$ , осуществляется итерационно с шагом  $\Delta \omega_j^{SG}$ .

## 2.4. Обучение классификатора

Организуем стохастический градиентный спуск по параметрам  $\omega(t) = \omega^{BP} + \omega^{SG}t$ . На каждой итерации будем производить один шаг градиентного спуска по параметрам  $\omega^{BP}$ , их обновленные значения использовать для вычисления шага градиентного спуска по параметрам  $\omega^{SG}$ , а затем обновлять значения  $\omega^{SG}$ . Также будем считать, что

$$\dot{y}_{kj} = \frac{y_{kj} - y_{kj-1}}{t}, \quad j = 2, \dots, M, \quad \dot{y}_{k1} = \dot{y}_{k2}, \quad k = 1, \dots, 2N.$$

Для остановки итерационного процесса обучения могут быть использованы различные методы. Чтобы избежать переобучения модели, остановка осуществляется в момент достижения функцией ошибки на тестовой выборке своего локального минимума.

## 2.5. Классификация

Элементы тестовой выборки обозначим  $\mathcal{Y}(t)$ :

$$\mathcal{Y}_{ij}(t) = \left( y_2(x_i(t)), \dots, y_M(x_i(t)), y(\hat{x}_j(t)) \right) \in \mathbb{R}^M, \quad i, j = 1, \dots, N.$$

Произведем классификацию элементов тестовой выборки. Пары  $(i, j)$ , определяющиеся из соотношения

$$\operatorname{argmax}_{(i,j)} \sum_{i=1}^N z(\mathcal{Y}_{ij}(t)), \quad (2.14)$$

задают взаимно-однозначное соответствие между наблюдениями  $y(X(t))$  и объектами  $X(t)$ , решающее поставленную задачу классификации.

## Глава 3

## Моделирование

Моделирование алгоритма выполнено на языке Python с использованием библиотеки машинного обучения PyTorch. В качестве прикладной задачи для моделирования рассматривается линейное движение объектов на плоскости (Рис. 3.1). Требуется классифицировать объекты на плоскости в момент времени  $t$  по предыдущим последовательным дискретным наблюдениям координат их местоположения.

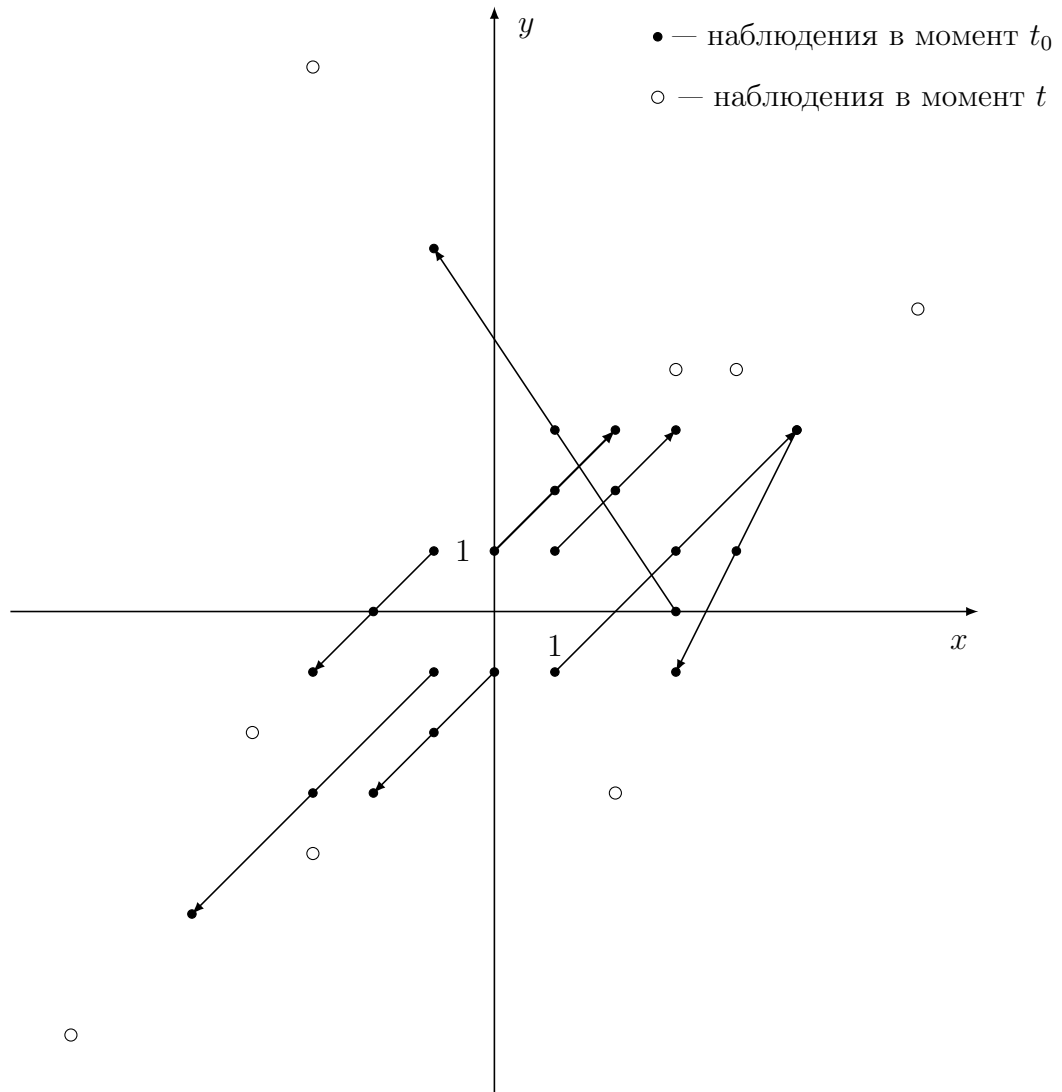


Рис. 3.1

Наилучшие результаты классификации были получены при следующих значениях параметров:

$$\dot{Q}_{\max} = -2, \quad \lambda = 0.001, \quad \gamma = 0.01.$$

Сравним полученные при моделировании графики функции ошибок  $Q(t)$  для методов скоростного градиента и обратного распространения ошибки.

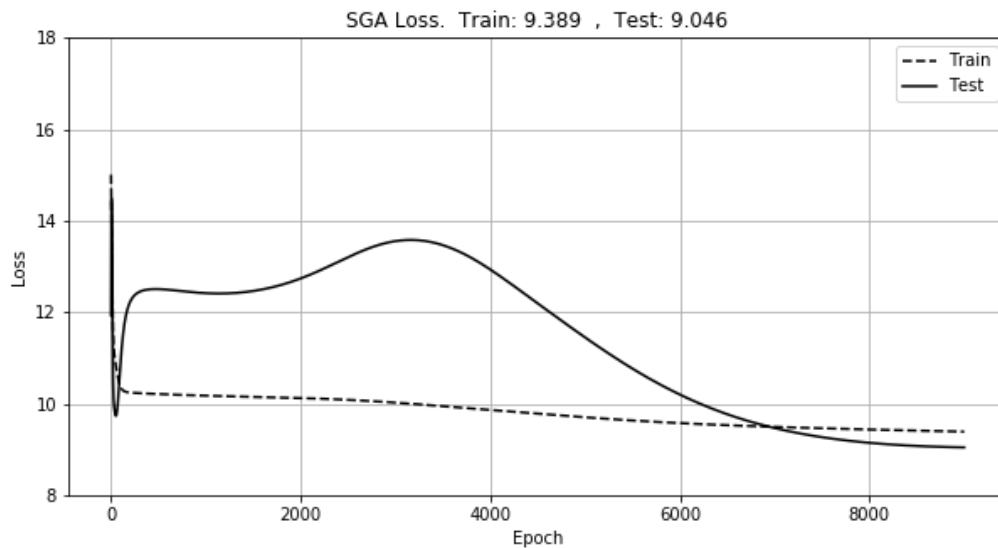


Рис. 3.2.  $Q(t)$  для алгоритма скоростного градиента

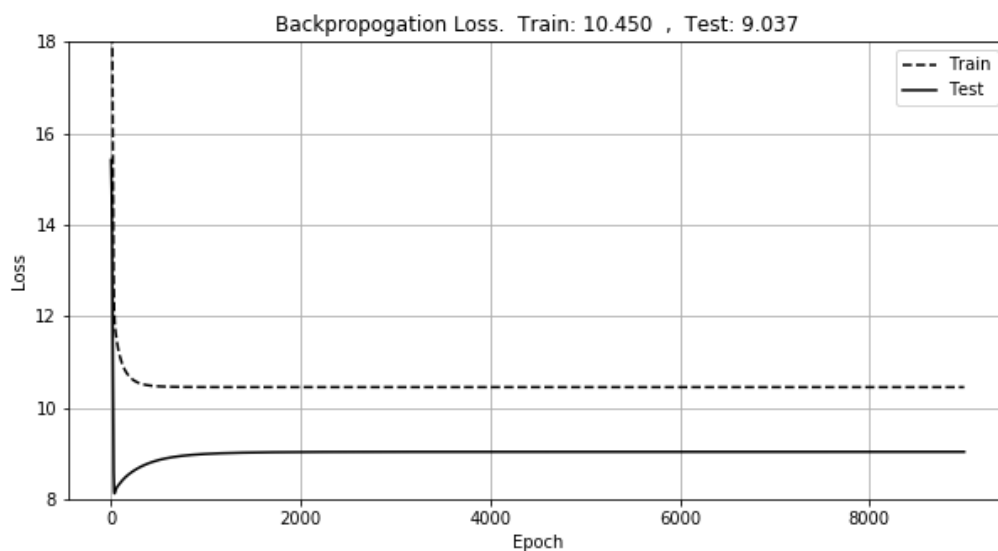


Рис. 3.3.  $Q(t)$  для метода обратного распространения ошибки

На приведенных графиках видно, что функция ошибок  $Q(t)$  на обучающей выборке достигает меньшего значения при обучении классификатора по алгоритму скоростного градиента. Значение функции ошибок  $Q(t)$  на тестовой выборке практически одинаковое, однако её минимум достигается значительно быстрее при обучении по методу обратного распространения ошибки. Сравним графики доли правильных ответов классификатора.

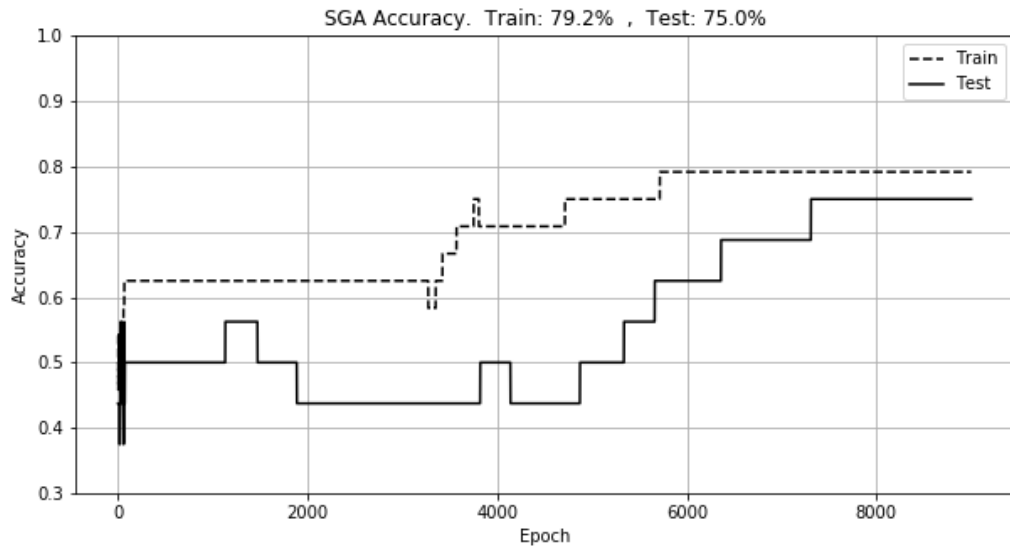


Рис. 3.4. Доля правильных ответов классификатора для алгоритма скоростного градиента

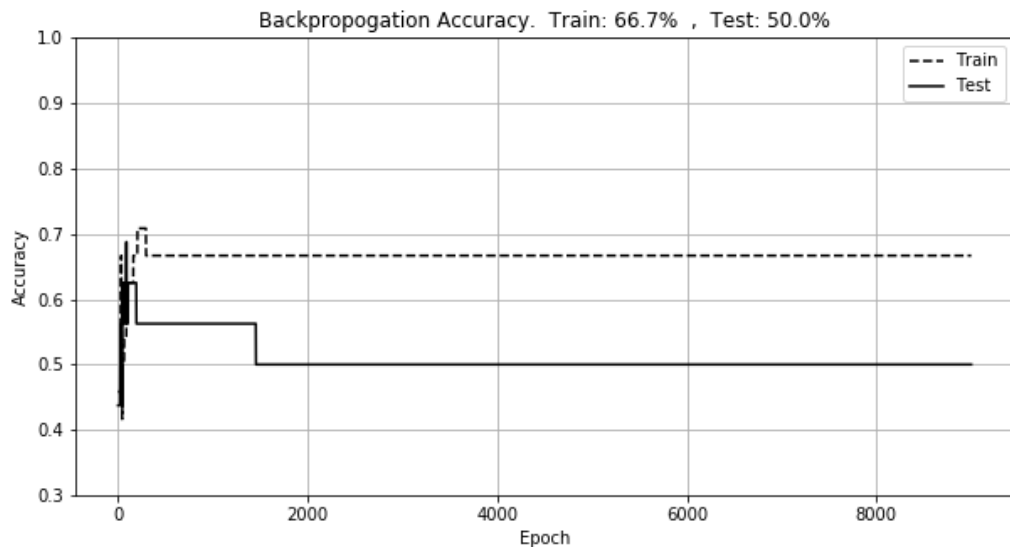


Рис. 3.5. Доля правильных ответов классификатора для метода обратного распространения ошибки

Перцептрон, обученный по алгоритму скоростного градиента, справляется с постав-

ленной задачей лучше на обучающей и на тестовой выборках. На тестовой выборке алгоритм скоростного градиента показал 75% правильных ответов, метод обратного распространения ошибки — 69% правильных ответов.

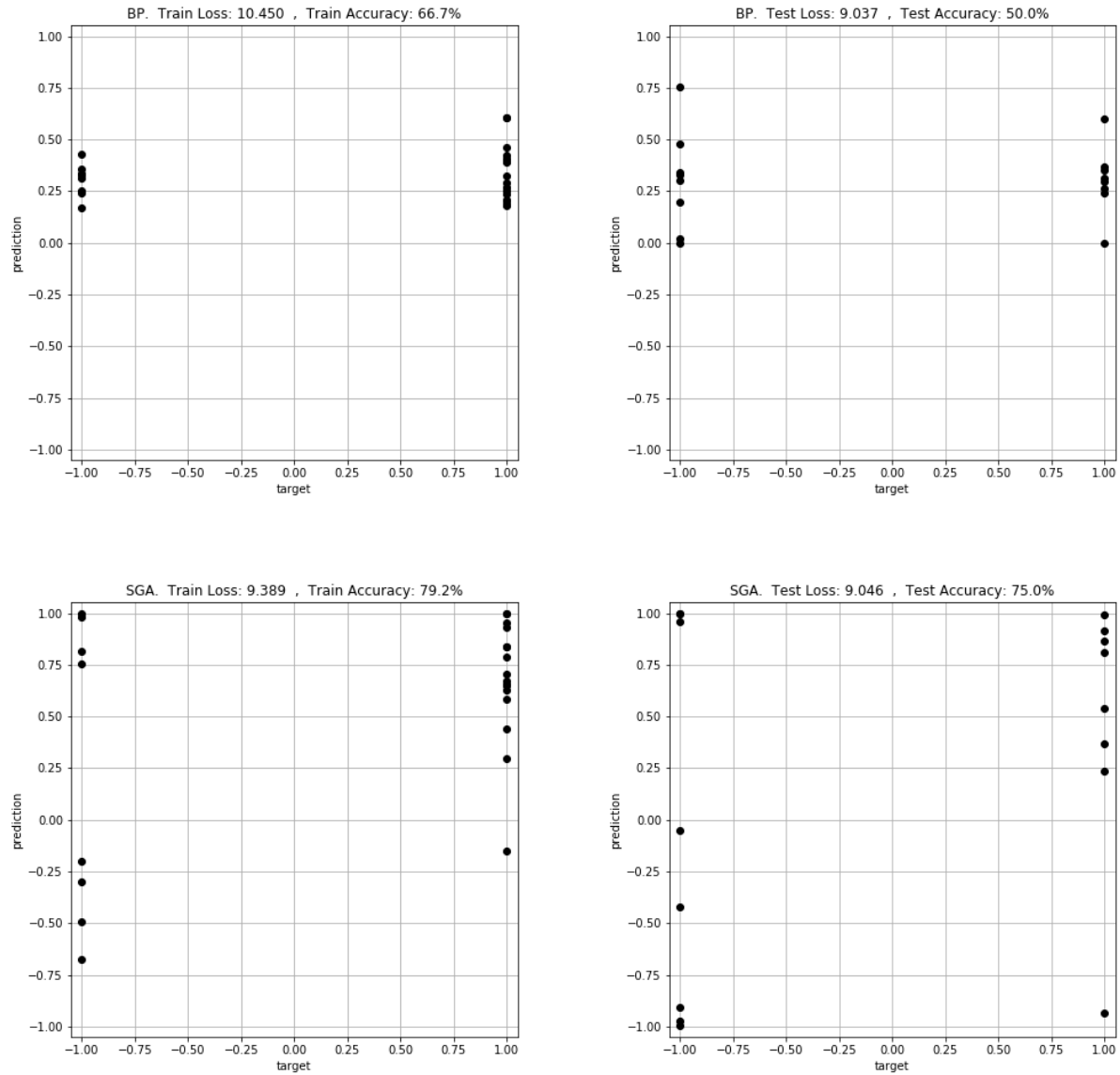


Рис. 3.6. Ответы классификаторов

На рисунке 3.6 продемонстрированы результаты работы перцептронов, обученных двумя методами. Показано, что перцептрон, обученный по методу обратного распространения ошибки, не смог разделить классы объектов.

## Заключение

Результатом данной работы является алгоритм классификации динамических объектов с помощью однослойного перцептрона, обученного по методу градиентной обратной связи. Результаты моделирования показали, что перцептрон с временной поправкой на веса  $\omega_j^{SGt}$  обеспечивает лучшее качество классификации, чем без этой поправки.

## Список литературы

1. Гелиг А.Х., Матвеев А.С., Введение в математическую теорию обучаемых распознающих систем и нейронных сетей. СПб.: Издательство СПбГУ, 2014.
2. Rumelhart D.E., Hinton G.E., Williams R.J., Learning Internal Representations by Error Propagation. Cambridge, MA: MIT Press, 1986.
3. Дорофеев А.С., Метод градиентной обратной связи в задаче обучения нейросетей. Материалы XXI конференции молодых ученых «Навигация и управление движением» с международным участием. СПб.: ГНЦ РФ АО «Концерн «ЦНИИ «Электроприбор», 2019. С. 310-311
4. Мирошник И.В., Никифоров В.О., Фрадков А.Л., Нелинейное и адаптивное управление сложными динамическими системами. СПб.: Наука, 2000.