

Исследование алгоритмов обучения искусственной нейронной сети для задач классификации

Корыстов Максим

студент 344 группы

Научный руководитель

Константин Невоструев

аспирант СП

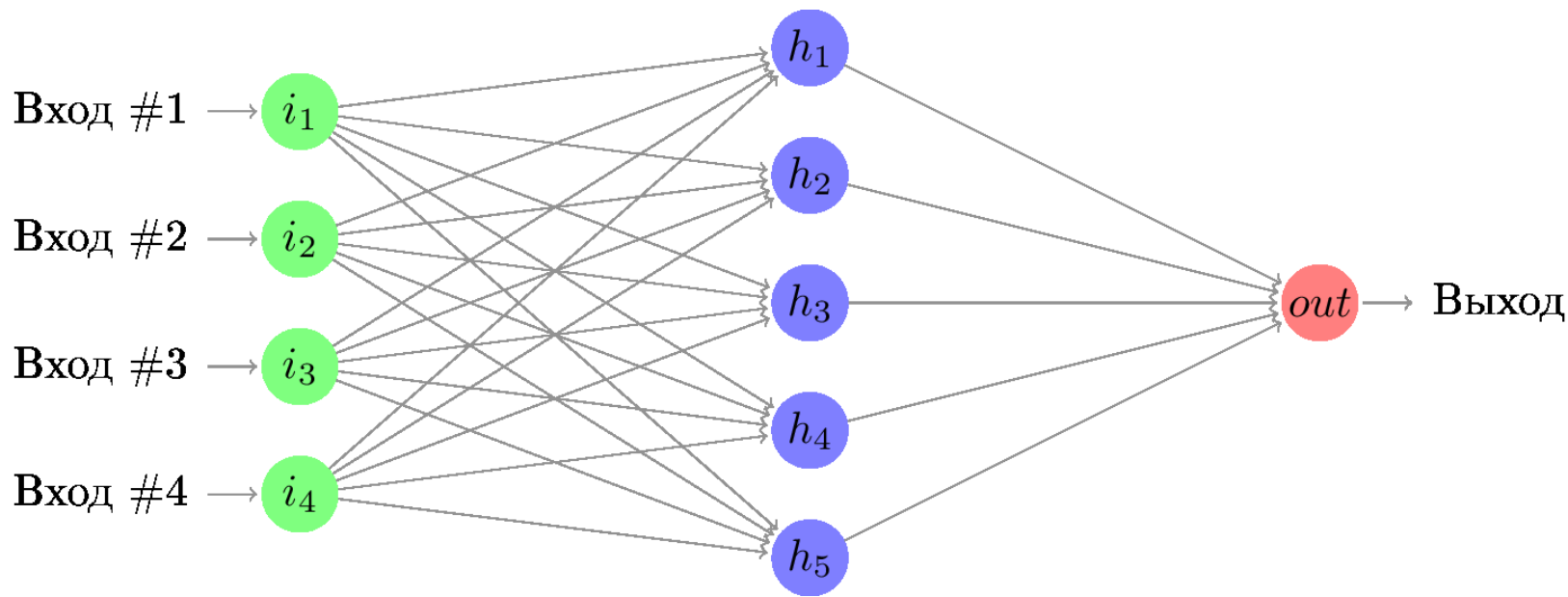
Задача классификации

Есть объекты с n свойствами разбитые на k классов. Классы известны только для m объектов. Надо найти

$$f : \mathbb{R}^n \rightarrow \{0, 1, \dots, k - 1\}$$

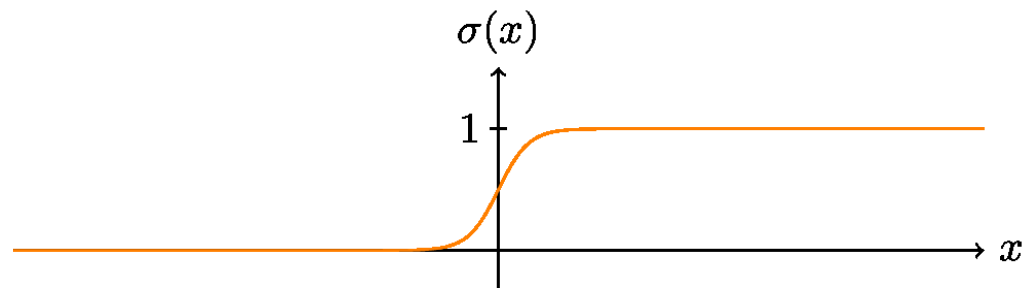
которое классифицирует все объекты.

Искусственная нейронная сеть



$$\bar{h} = \sigma(\Theta_1 \bar{i}), \quad out = \sigma(\Theta_2 \bar{h})$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



Цели работы

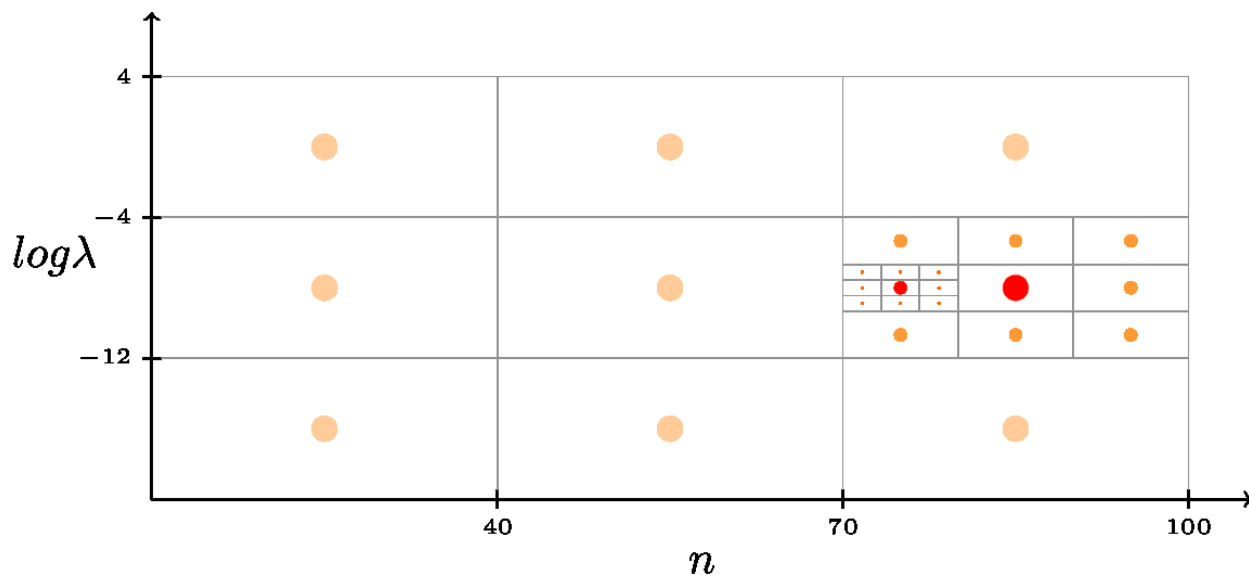
1. Подготовить наборы данных для задачи классификации
2. Произвести сравнение результатов алгоритмов обучения на разных наборах данных
3. Установить зависимость эффективности обучения нейронной сети использующую разные алгоритмы обучения от набора входных данных

Этап I. Автоматический выбор параметров

Параметры:

1. Количество нейронов в скрытом слое
2. Параметр переобучения

Поиск в сетке (Grid Search)



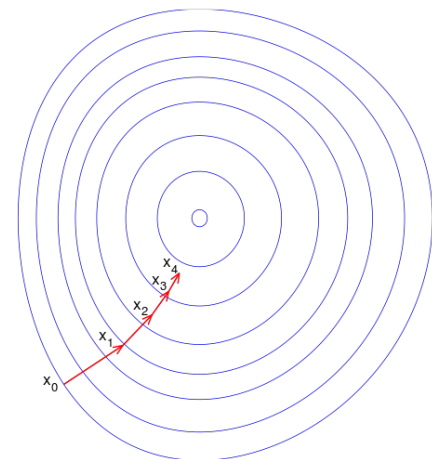
Этап II. Обучение искусственной нейронной сети

Функция штрафа $J(\Theta)$

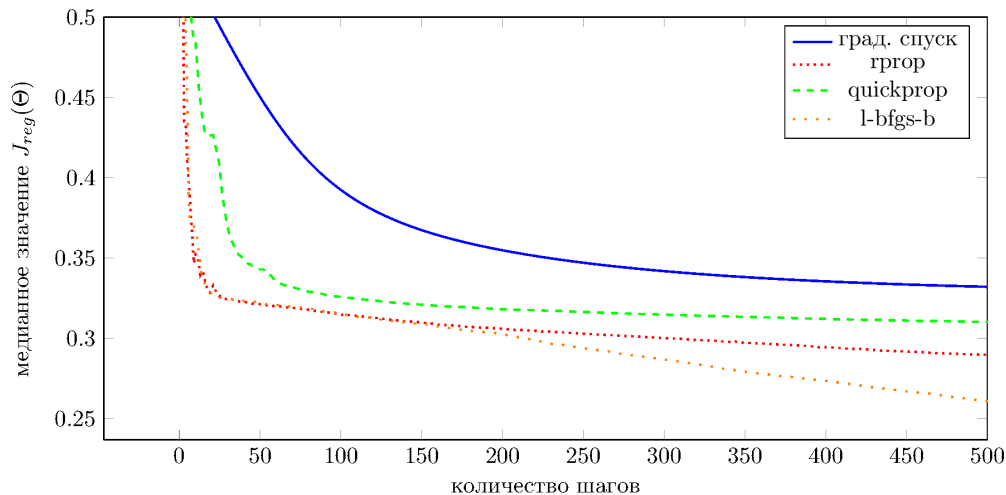
Можно найти $\text{grad } J(\Theta)$ методом обратного распространения ошибки (*backpropagation*).

Поиск минимума функции штрафа - улучшение классификатора.

Алгоритмы: *Градиентный спуск*, *RPROP*, *QuickProp*, *L-BFGS-B*.



Сравнения 1



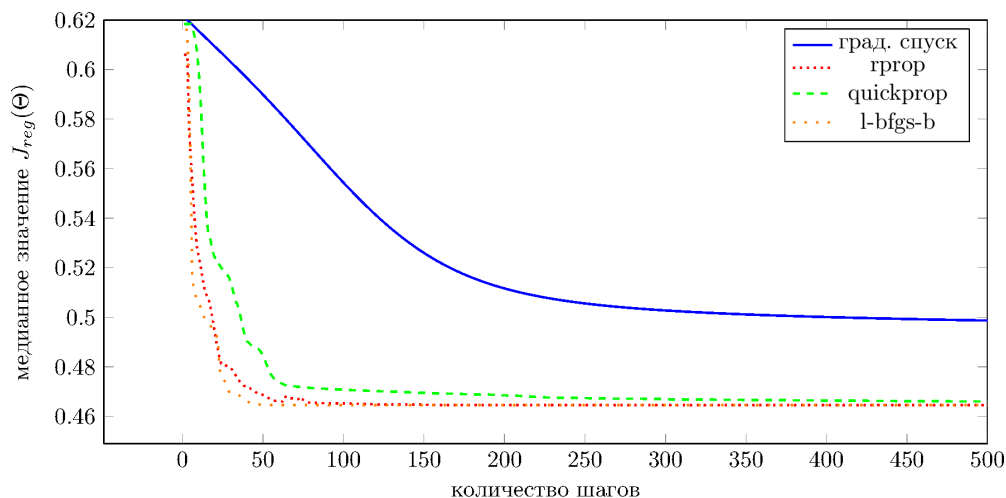
Набор данных adult

тренировочная выборка: 32 561

тестовая выборка: 16 281

атрибуты у объекта: 123

Бинарная задача классификации



Набор данных titanic

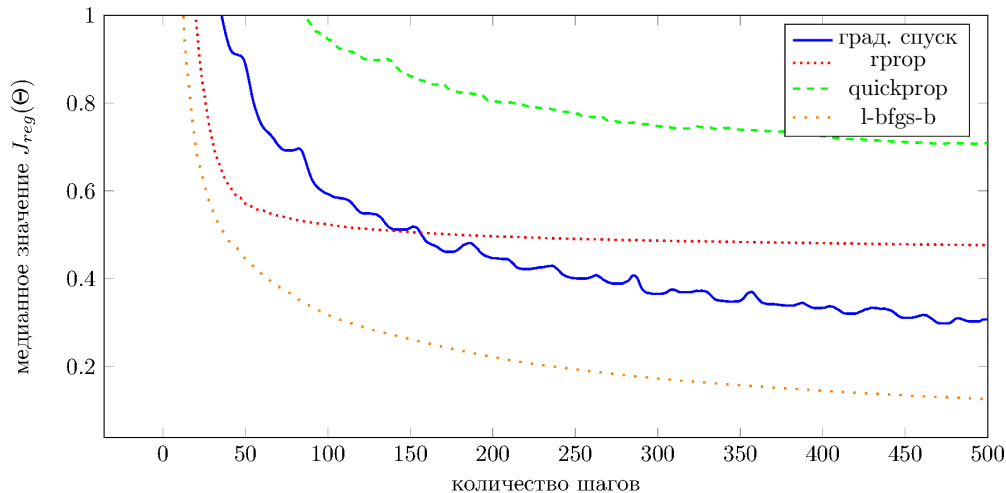
тренировочная выборка: 1526

тестовая выборка: 673

атрибутов у объекта: 6

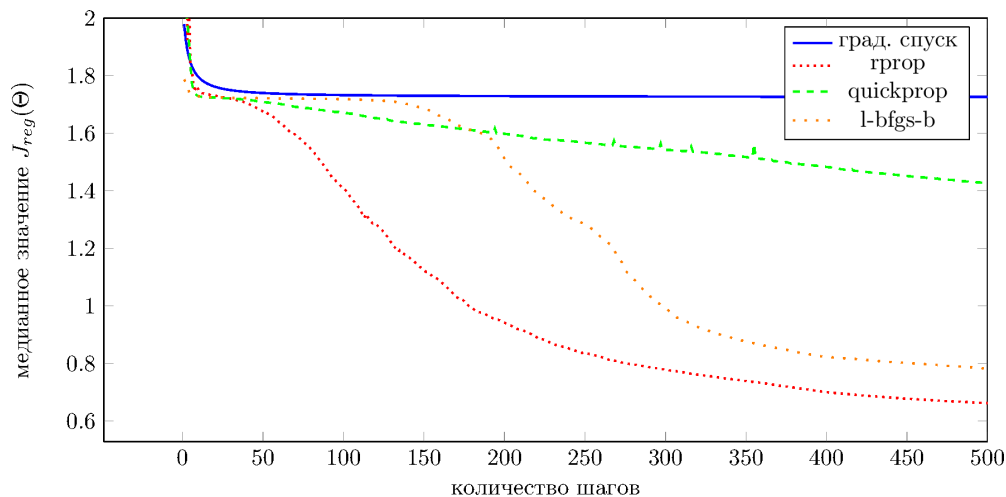
Бинарная задача классификации

Сравнения 2



Набор данных mnist

тренировочная выборка: 60 000
 тестовая выборка: 10 000
 атрибуты у объекта: 784
 10 классов



Набор данных poker

тренировочная выборка: 25 010
 тестовая выборка: 1 000 000
 атрибутов у объекта: 17
 10 классов

Финальное сравнение точности

	Наборы данных							
	adult		titanic		mnist		poker	
Алгоритмы	%	F1	%	F1	%	F1	%	F1
Град. спуск	82.85	63.89	77.22	50.91	94.07	93.99	50.04	6.75
RPROP	83.44	64.30	77.31	52.11	90.64	90.51	92.43	19.91
QuickProp	83.06	64.01	77.31	52.11	88.85	88.67	69.05	14.14
L-BFGS-B	84.05	64.58	77.31	52.11	95.97	95.93	91.02	18.88

F1 - оценка точности классификатора, значения от 0 до 1, больше - лучше

Результаты

- Произведены сравнения алгоритмов Градиентный спуск, RPROP, QuickProp, L-BFGS-B на различных наборах данных
- Выявлена связь между характеристиками наборов данных и эффективностью различных алгоритмов обучения

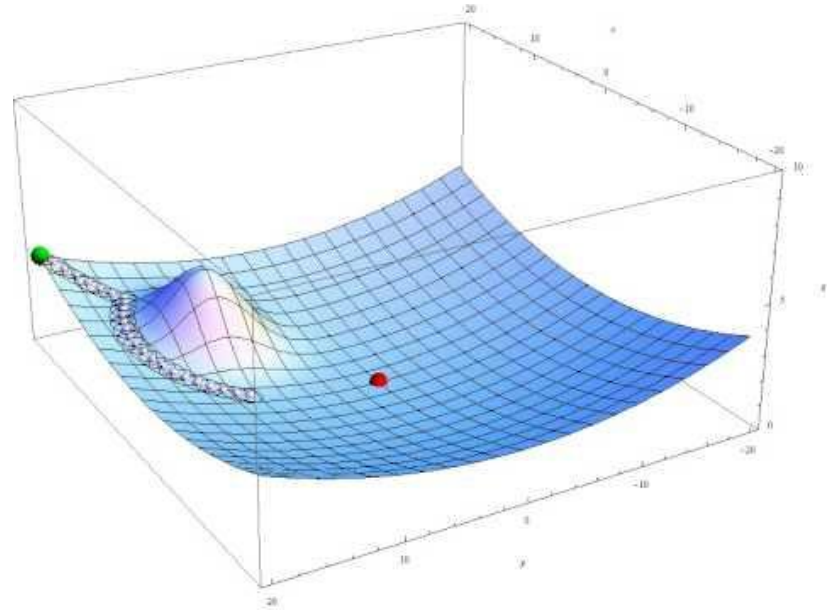
Дополнительный слайд 1

Функция штрафа

$$J(\Theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=0}^{K-1} y_k^{(i)} \log(h_{\Theta}(x^{(i)})_k) + (1 - y_k^{(i)}) \log(1 - (h_{\Theta}(x^{(i)}))_k) \right] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} \Theta_{ji}^{(l)2}$$

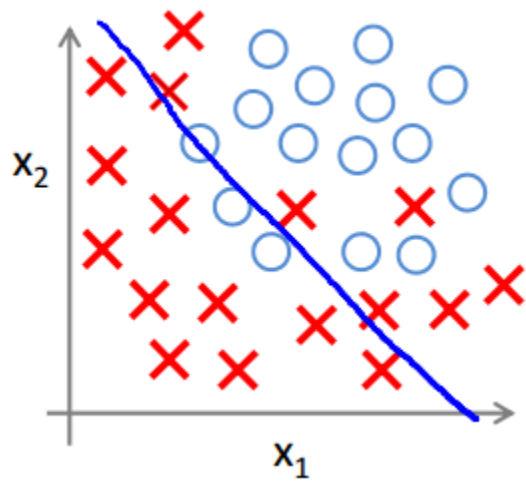
Градиентный спуск

$$\theta_i = \theta_i - \alpha \frac{\partial J(\Theta)}{\partial \theta_i}$$

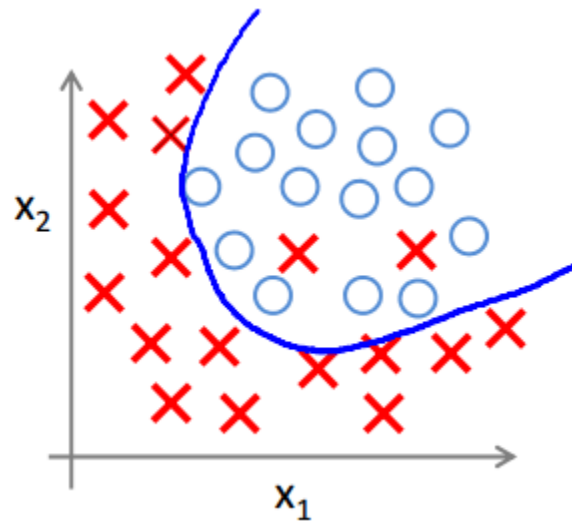


Дополнительный слайд 2

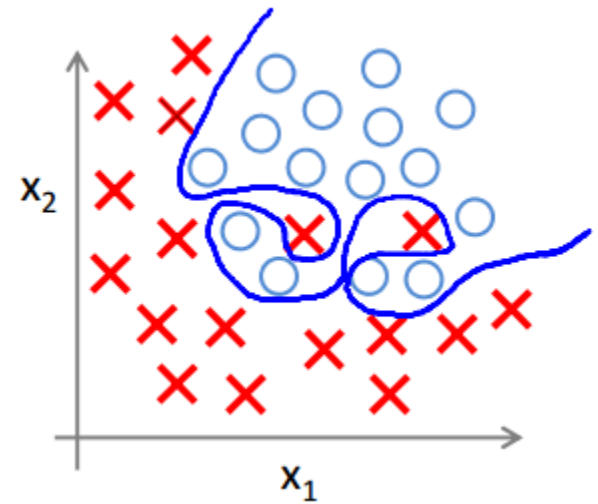
Недообучение



Удв. обучение



Переобучение



Дополнительный слайд 3

MNIST (Рукописные цифры) [60,000; 10,000]

Алгоритм	Точность (%)	F1 * 100	Время обучения (с)
Нейронная сеть	96.39	96.34	911
SVM SMO	98.57	-	3578
ELM	96.10	96.00	807
Лучший	99.77	-	-

Рокер (Классифицировать руку) [25,010; 1,000,000]

Алгоритм	Точность (%)	F1 * 100	Время обучения (с)
Нейронная сеть	92.49	20.94	246
SVM SMO	92.38	-	4200
ELM	72.5	69	807

F1 - оценка точности классификатора, значения от 0 до 1, больше - лучше

Дополнительный слайд 4

Adult (Предсказание дохода) [32,561; 16,281]

Алгоритм	Точность (%)	F1 * 100	Время обучения (с)
Нейронная сеть	85.08	65.69	79
SVM SMO	85.08	-	197
ELM	85.00	65.00	39

Titanic (Выживание пассажира) [1,526; 673]

Алгоритм	Точность (%)	F1 * 100	Время обучения (с)
Нейронная сеть	77.62	60.91	1
SVM SMO	77.31	-	1
ELM	77.30	52	1

Обучение производилось на компьютерах с разной вычислительной мощностью