

Сравнение и тестирование реализаций алгоритма SVM для задач классификации



курсовая работа
студента 445 группы
Дудина Виктора

научный руководитель
Константин Невоструев
аспирант кафедры СП

СПбГУ
2013

Постановка задачи классификации



Имеется множество объектов, разделённых на классы

Каждый объект описывается набором определенных параметров

Задано конечное множество объектов, для которых известно, к каким классам они относятся

Классовая принадлежность остальных объектов не известна

Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества на основе его параметров

Цели работы



Реализовать различные алгоритмы обучения Support Vector Machines (SVM) для задач классификации

Подготовить данные для проведения тестирования

Провести сравнительное тестирование качества и скорости для алгоритмов обучения SVM

Провести сравнительное тестирование SVM и других алгоритмов, решающих задачу классификации

Исследованные алгоритмы



Реализованные алгоритмы SVM:

- **Sequential Minimal Optimization (SMO) with Kernel Function**
- **Sequential Minimal Optimization (SMO) with Linear Separator**
- **Active Set Method (ASM)**

Сторонние алгоритмы:

- **Artificial Neural Network (Encog)**
- **Decision Tree (jaDTi)**

Результаты



Реализованы все описанные алгоритмы SVM, получены сторонние библиотеки с другими алгоритмами

Найдены и подготовлены данные для проведения тестирования

Создана автоматизированная система для проведения тестирования

Проведены тесты

Тестирование. “Telescope”



10 параметров, 3400 элементов в обучающей выборке

Таблица 1: Результаты для набора Telescope

	ASM	SMO
Точность обучения, %	86.16	86.17
Время обучения, сек	123	141
Время на 1 предсказание, мсек	0.1809	0.1770

Тестирование. “Spambase”



57 параметров, 1250 элементов в обучающей выборке

Таблица 2: Результаты для набора Spambase

	ASM	SMO
Точность обучения, %	92.03	92.03
Время обучения, сек	8	49
Время на 1 предсказание, мсек	0.0902	0.0889

Тестирование. “Musk”



166 параметров, 1450 элементов в обучающей выборке

Таблица 3: Результаты для набора Musk

	ASM	SMO
Точность обучения, %	97.49	97.49
Время обучения, сек	7	83
Время на 1 предсказание, мсек	0.1964	0.1977

Тестирование. Деревья решений



Таблица 4: Результаты тестирования Деревя Решений

	Telescope	Spambase	Musk
Точность обучения, %	83.85	89.48	93.26
Время обучения, сек	< 0.5	< 0.5	1
Время на 1 предсказание, мсек	0.0005	0.0010	0.0027

Тестирование. Нейронные сети



Таблица 5: Результаты тестирования **Нейронной сети**

	Telescope	Spambase	Musk
Точность обучения, %	85.09	91.72	94.80
Время обучения, сек	12	3	2
Время на 1 предсказание, мсек	0.0054	0.0105	0.0062