

# **Исследование и сравнение алгоритмов SVM и TSVM для задач классификации**

Гарифуллин Шамиль, 344 гр.  
Руководитель: Невоструев Константин

СПбГУ  
2014

# Постановка задачи классификации

Дано:

- фиксированное дискретное множество классов
- множество данных с фиксированными параметрами

Необходимо:

- построить отображение, сопоставляющее данным их классы, используя
  - маркированные данные (SVM)
  - как маркированные, так и немаркированные данные (TSVM)

# Цели работы

- Реализовать алгоритмы SVM и TSVM для задач классификации
- Подготовить данные для проведения тестирования
- Провести сравнение качества и скорости работы алгоритмов обучения SVM и TSVM

# Данные “Diabetes”

8 параметров. 576 элементов на обучение,  
192 на проверку

	SVM	TSVM
Точность обучения	0,75±0.02	0,67±0.02
F1-score	0,61±0.01	0,65±0.01
Время на обучение, с.	0,43±0.12	3.5±0.57

# Данные “Titanic”

6 параметров. 1526 элементов на обучение,  
673 на проверку

	SVM	TSVM
Точность обучения	$0,77 \pm 0.02$	$0,51 \pm 0.03$
F1-score	$0,57 \pm 0.01$	$0,61 \pm 0.01$
Время на обучение, с.	$0,96 \pm 0.07$	$25,2 \pm 3.21$

# Данные “Reuters”

9947 параметров. 10 элементов на обучение,  
600 на проверку

	SVM	TSVM
Точность обучения	0,84±0.03	0,96±0.01
F1-score	0,79±0.02	0,92±0.01
Время на обучение, с.	0,45±0.15	16±2.42

# Итоги

- Реализованы описанные алгоритмы
- Найдены и подготовлены данные для проведения тестирования
- Проведены тесты
- Получена сравнительная оценка работы алгоритмов