

Замечания к программной реализации задач осеннего семестра и графику консультанта

Ю.В. Максимов,
ИППИ РАН, ПреМоЛаб МФТИ
yury.maximov@phystech.edu, yuri.maximov@inria.fr

4 сентября 2014 г.

Задачи. В первую очередь относиться к задачам 8,9,25,38,39.

Консультант. По ряду задач консультации будут как с Сергеем Грудининым так и с Юрием Максимовым. Последний доступен по скайп `yuramaximov`, и указанным в шапке электронным адресам. Доступен по скайп/е-mail для обсуждения любых тем, связанных с задачами по субботам, начиная с 17 МСК, по вторникам начиная с 20 МСК. В остальное время – нерегулярно.

Описание. В предлагаемых задачах особое внимание уделяется применению методов выпуклой и невыпуклой оптимизации к решению задач анализа данных. Многие из рассматриваемых в работе алгоритмов достаточно тяжелые (например имеют кубическую сложность по размеру пространства переменных). Это делает их сложно реализуемыми на рабочих машинах.

Математические методы “ускорения”. Снижение размерности входных данных. В первую очередь это PCA, см. также вариант построения случайных проекций (работа М.Е. Карасикова в предыдущем семестре). Как правило датасеты достаточно большие, так что снижением размерности рекомендуется не пренебрегать.

Программные методы “ускорения”. В настоящем семестре работа по задачам ограничивается прототипированием. Для этого можно использовать MATLAB, однако использование C/C++ не может не приветствоваться. Для программной реализации части задачи рекомендуется использовать современные библиотеки анализа данных, такие как VLFEAT и YAEL(опционально), которые в настоящее время поддерживаются лишь для Linux 64 bit и Mac OS X 64 bit. Использование этих библиотек позволяет существенно ускорить и повысить качество решения задач. Однако оно не является обязательным, для реализации основных программных компонентов достаточно библиотек LIBSVM(от которой нам будет нужен модуль загрузки данных) и SVMLIN, предоставляющий относительно функционал для обучения линейных классификаторов в задачах полного и частичного обучения большой размерности. И LIBSVM и SVMLIN можно использовать под Windows 32/64. Ссылки на библиотеки, представлены в параграфе “Литература”. Для решения выпуклых задач оптимизации рекомендуется использовать CVX в комбинации с солвером MOSEK. Все библиотеки являются свободно распространяемыми. В случае трудностей с их установкой и использованием, обращайтесь к консультанту.

1. Библиотека VLFEAT;
2. Библиотека LIBSVM;
3. Библиотека SVMLIN;
4. CVX является частью Матлаба. Вместе с CVX лучше всего использовать Солвер MOSEK, документация по которому доступна тут и тут.