

Scalable Linear Models

В этом проекте необходимо реализовать систему для обучения линейных моделей – быстро работающую и поддерживающую, как минимум, такие наборы букв : sgd, ftrl-proximal, adagrad. Утилита должна уметь сохранять обученную модель в удобном формате, и уметь применять её. Классификацию и регрессию нужно поддерживать в виде отдельных режимов.

Интерфейс

Утилита должна иметь два режима - обучение и применение. Работать нужно со стандартными файлами csv. При обучении она должна принимать, как минимум, путь к файлу с данными, тип обучаемого алгоритма, максимальный допустимый размер модели и темп обучения.

Принцип работы

Хочется, чтобы утилита была быстра и качественна. Постарайтесь применить все описанные на лекциях трюки, хешируйте фишки, следите за эффективностью работы с памятью. Постарайтесь быть не медленнее чем vowpal wabbit.

Процесс сдачи задания

Проект рассчитан на 1-3 человек.

Реализуйте библиотеку, так как написано выше. Сделать это надо на любом компилируемом языке, позаботившись о качестве кода.

Библиотеку разместите на github. Напишите к ней приятный readme, в котором будет общее описание библиотеки и инструкция о том как её запустить и использовать.

Сравните свою библиотеку с популярными аналогами ? vowpal wabbit и liblinear. Оцените скорость работы и качество полученной модели на датасетах Criteo и avazu. Опирайтесь в бенчмарках можете на этот пост. О результатах бенчмарков также напишите в readme вашего репозитория.