Машинное Обучение МФТИ
Практическое задание №3: Линейные модели

17 апреля 2016 г.

Введение

В 3 практическом задании вы подробно изучите то, как работают линейные модели. Целью задания является решение задачи классификации текстов, с использованием логистической регрессии и алгоритма опорных векторов, который вы реализуете самостоятельно.

Настройка окружения

Для успешного решения практического задания, на вашем компьютере должна быть установлена Anaconda, скачать ее можно тут: `https://www.continuum.io/downloads`

Сами задания находятся внутри файлов linear_classifier.py linear_svm.py softmax.py и ноутбука practical_3.ipynb

Краткое описание задачи

Вам предстоит решить задачу классификации на датасете Amazon Fine Foods Reviews. Описание датасета и задачи можно посмотреть тут: `https://www.kaggle.com/snap/amazon-fine-food-reviews` Чтобы упростить задачу мы будем использовать только summary отзыва, без основного текста. Данные можно скачать на странице соревнования.

Шаги решения и отчет

Шаги решения описаны в основном ноутбуке. В качестве отчета необходимо предоставить файлы с моделями linear_classifier.py linear_svm.py softmax.py и ноутбук practical_3.ipynb. Также необходимо получить хороший скор в соревновании Kaggle In Class.

Методические указания

1. При подборе параметров модели рекомендуется использовать только часть обучающей выборки, для того чтобы сократить время обучения.

2. Согласно правилам соревнований нельзя делать больше 3х коммитов в систему в сутки. Из этого надо сделать следующие выводы:

(a) Обучаться нужно локально (cross-validation) и только после получения результата, который вы считаете удовлетворительным, нужно делать submit в систему.

(b)Начать делать домашнее задание стоит заблаговременно.

3. Обратите внимание, что публичные результаты на kaggle рассчитываются только по части контрольной выборки, и будут рассчитаны по всей контрольной выборке после окончания соревнования. Будьте аккуратны с переобучением.

4. Победители получают бонусные балы – шарить решение не выгодно.

Разница между списыванием и помощью товарища иногда едва различима. Мы искренне надеемся, что при любых сложностях вы можете обратиться к семинаристам и с их подсказками самостоятельно справиться с заданием. При зафиксированных случаях списывания (одинаковый код, решение задачи), баллы за задание будут обнулены всем участникам инцидента.