

СЕССИЯ 3

Исходные файлы:

- | | |
|-------------------------------|--------------------------------|
| 1) sample.csv | (Пример файла прогнозов) |
| 2) Data.zip | (Результаты предыдущей сессии) |
| 3) Машинное обучение – C3.pdf | (Инструкция к 3 сессии) |

Результаты работы:

- | | |
|------------------------------------|------------------------------|
| 1) Data.zip | (Предобработанные данные) |
| 2) Report_C3.html, Report_C3.ipynb | (Отчет о проделанной работе) |
| 3) Readme.txt | (Дополнительные комментарии) |

ВВЕДЕНИЕ

В этой сессии вы продолжаете работать с данными, подготовленными в предыдущей сессии. Необходимо выполнить обучение нескольких моделей классификации, выбрать более точную и осуществить для нее настройку гиперпараметров.

Какая-либо работа, обусловленная задачами предыдущей сессии, выполненная в ходе текущей, оцениваться не будет, поэтому проделывайте её только в случае необходимости.

ЗАДАНИЕ

3.1 Сравнение нескольких алгоритмов на сессиях

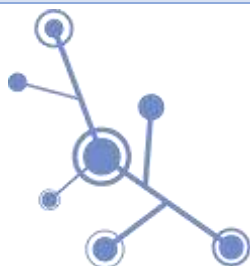
Выполните разбиение выборки на обучающую и отложенную. Проведите обучение моделей, основанных на различных алгоритмах. Сделайте прогноз в виде предсказанных вероятностей того, что сессия является сессией злоумышленника на отложенной выборке. Выполните оценку моделей разной степени сложности в соответствии с выбранной метрикой, предусмотрите распараллеливание процессов. Выберите модель, показавшую лучшее качество. Дополните набор данных признаками, полученными в предыдущей сессии, чтобы добиться более точной работы выбранной модели.

3.2 Гиперпараметрическая оптимизация модели

После выбора модели необходимо оптимизировать её под решаемую задачу, настраивая гиперпараметры. Построить для данной модели кривую валидации (зависимость качества классификации от одного из гиперпараметров алгоритма) и кривую обучения (зависимость качества классификации от объема выборки). Выполните оценку данной модели после настройки гиперпараметров.

3.3 Прогноз классификации для тестовой выборки

Сделайте прогноз в виде предсказанных вероятностей отнесения к классу 1 для тестовой выборки с помощью полученной модели, обученной уже на всей обучающей выборке. Запишите ответы в файл в соответствии с шаблоном sample_sub.csv



3.4 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. Архив Data.zip должен содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов архива Data.zip.

