

Анализ данных в бизнесе 2021

Домашнее задание №2

Дедлайн: 04.03.2021 23:59 (Moscow time)

Alexey Romanenko, Alexey.Romanenko@sas.com

Организационные требования

Решения присылать через форму приёма ДЗ <https://forms.gle/bzf5ksDNCG77cuo4A>.

Требования к оформлению вложений см. ниже.

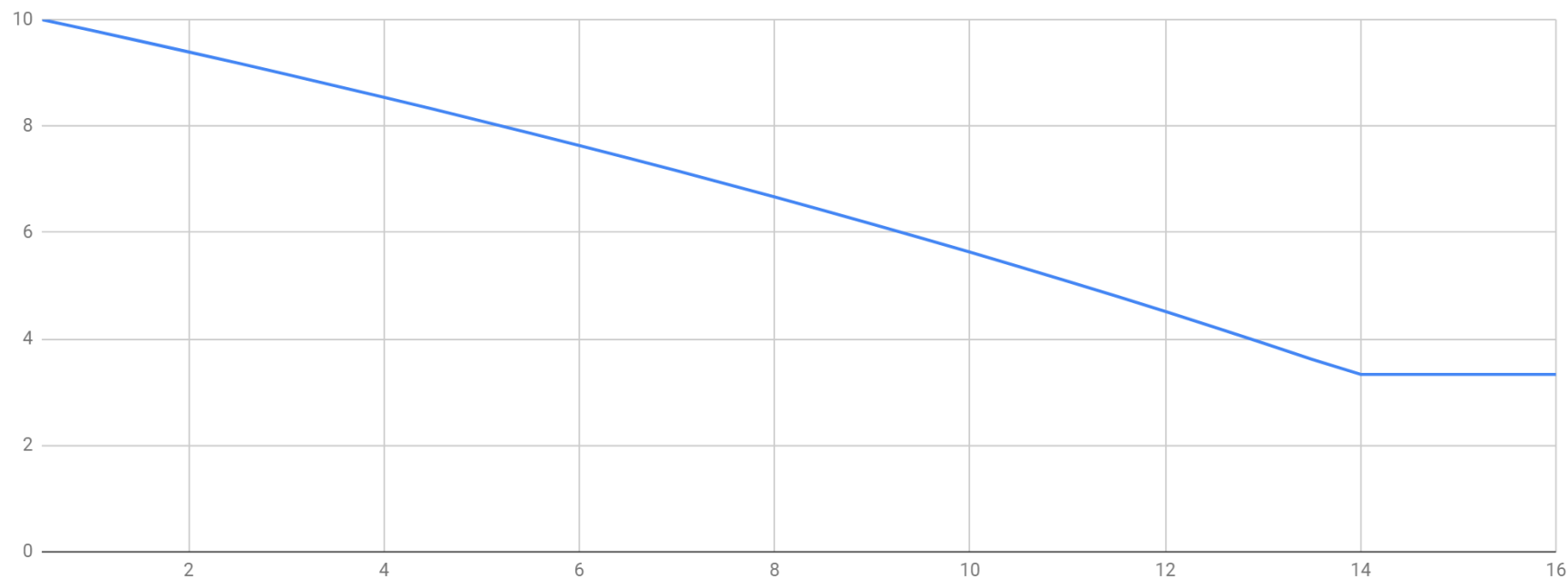
Решения, присланные не в соответствии с требованиями в форме, проверяться не будут.

Решения, присланные после даты дедлайна, будут оштрафованы: максимальный балл за ДЗ будет снижаться в зависимости от величины просрочки дедлайна (в днях):

$$\text{МаксимальныйБалл} = \max\left(\frac{1}{3}, \log_2\left(1.23 + 0.77 * \max\left(0, \left(1 - \frac{\text{ДнейПослеДедлайна}}{14}\right)\right)\right)\right).$$

График зависимости максимально возможного балла за ДЗ от просрочки дедлайна приведён ниже:

Максимальный балл за ДЗ после просрочки дедлайна



Описание задание

1. Требуется принять участие в конкурсе по прогнозированию спроса на новые пары товар|магазин:

<https://www.kaggle.com/t/d6212c14bc11462d996f0643bdfef68f>

- а. Формат csv-файла:

id, predicted

1, 2

1, 3

1, 0

2. Можно использовать любые реализации изученных на курсе моделей машинного обучения и не только их.

Критерии оценки (максимум 10 баллов)

1. Для того чтобы получить оценку, требуется прислать код (python/R/sas) или проект (в SAS VDMML), который генерирует финальный сабмит (лучший на Public из тех, что вы выбрали в качестве финальных), файл требуется отправить в [форму](#).
2. Название файла требуется отправлять строго по шаблону <Name>_<Surname>_<groupNo>_hw2.csv, например, Ivan_Ivanov_41_hw2.csv).
3. Результат за решения контеста измеряется от 0 до 10:
 - а. Если результат превосходит (на Private) результат Benchmark_1, тогда оценка будет 8 или более
 - б. Если результат превосходит (на Private) результат Benchmark_2, тогда оценка будет 10.
 - с. Остальная градация оценок будет сформирована по результатам контеста в зависимости от результатов всех участников.
4. Два одинаковых или почти одинаковых присланных кода будут оштрафованы: 0 за задание независимо от результатов в Leaderboard.