

Решение заданий направления Risk Data Scientist, ЮниКредит Банк**Пример решения задания 1 от эксперта****Задача 1**

Ответ: среднее. В случае константного предсказания метрика имеет вид $MSE = \frac{1}{N} \sum_{n=1}^N (\hat{y} - y_n)^2$. Продифференцируем ее один раз по \hat{y} и приравняем производную нулю $\frac{\partial MSE}{\partial \hat{y}} = 0$. Отсюда получаем, что $\hat{y} = \frac{1}{N} \sum_{n=1}^N y_n$. При повторном дифференцировании убеждаемся, что на таком \hat{y} достигается минимум $\frac{\partial^2 MSE}{\partial^2 \hat{y}} = 2 > 0$.

Задача 2

Ответ: никак. Метрика GINI служит оценкой качества ранжирования наблюдений по вероятности принадлежности классу с реализованным целевым событием (например, по вероятности выхода заемщика в дефолт). Взятие монотонной функции (например, кубической) от предсказания не меняет порядок ранжирования.



Пример решения задания 2 от эксперта

Скрипт можно скачать, перейдя по ссылке:

<https://drive.google.com/file/d/1-lRDPzAgyDbPmFlklJUzBib0d5ByoEfc/view?usp=sharing>



Пример решения задания 3 от эксперта

Скрипт можно скачать, перейдя по ссылке:

https://drive.google.com/file/d/10pXvVBuXi3sYhOTpVsc0l3qfDUZJ_LP0/view?usp=sharing

Решение находится в файле «Task3_greedy_selection.py». У него более широкая реализация по сравнению с поставленной задачей – его можно считать своего рода вознаграждением за ваши усилия. Можно применять его для любых своих целей в дальнейшем. Успехов!