

Соревнование по курсу «Байесовский выбор моделей»

Общая информация

- Участие в соревновании по командам. В каждой команде от 3 до 5 человек;
- Каждый участник победившей команды получает 150 баллов, баллы последующих участвовавших команд убывают до 20 от 100 баллов для второй;
- Первый этап состоит в анализе подготовленных временных рядов (4 набора);
- Результаты на данных первого этапа нужно отправить на почту aduenko1@gmail.com и iakovlev.kd@phystech.edu до 6 апреля 23:59 по Москве;
- После проверки результатов первого этапа командам будут доступны правильные ответы для разладки каждого ряда, а также прогнозы и распределения вкладов команд противников для калибровки финального решения;
- Финальное решение в виде кода будет тестироваться на разных выборках.

Описание процедуры соревнования

Прогноз временного ряда. Имеется временной ряд $\mathbf{x}_1, \dots, \mathbf{x}_N$, $\mathbf{x}_i \in \mathbb{R}^D$, сгенерированный по некоторому правилу. При этом в этом правиле могла произойти разладка в некоторый момент времени T .

Задача: Найти момент разладки $100 \leq T \leq N - 100$ или указать, что разладки не было $T = N + 1$.

Условие победы: Побеждает та команда, прогноз $\hat{T}$ которой окажется ближайшим к истинному моменту разладки T .

Описание игры:

- Перед началом игры у каждой команды есть $S_0 = 1000000$ конфет;
- Игра состоит из поочередного анализа K временных рядов, для каждого из которых требуется построить прогноз $\hat{T}_k$;
- Каждая команда должна выбрать размер вклада D_k для каждого временного ряда, $\sum_k D_k \leq S$;
- Вклад победившей команды удваивается ($S_k = S_{k-1} + D_k$), проигравшая – теряет все конфеты ($S_k = S_{k-1} - D_k$). Если команд Z больше 2, то команды, занявшие места z со второго до предпоследнего, получают $S_k = S_{k-1} + D_k(1 - 2\frac{z-1}{Z-1})$ конфет;
- Переход на шаг $k + 1$.

Замечание 1: Требуется предоставить вклады D_k , а также метки T_k для временных рядов, предоставленных для анализа заранее (обучающая выборка). Формат ответа:

- Таблица с именем answers_partN.xlsx с индексом N , соответствующим номеру ряда в каждой выборке TS_N, и двумя колонками T и D .

Замечание 2: Кроме того, требуется предоставить алгоритм выбора D_k , T_k для временных рядов из контрольной выборки (количество то же, что и на обучении).

Замечание 3: Алгоритм должен быть выполнен на питоне в виде функции с заданным интерфейсом и предобучен по данным из обучающей выборки. Интерфейс будет сообщен

дополнительно после анализа результатов на предоставленных выборках.

Замечание 4: Время работы алгоритма на всей контрольной выборке (равнозначна обучающей) - не более 30 минут на среднем ноутбуке.