

Соревнование по курсу «Байесовский выбор моделей»

Общая информация

- Заочная часть соревнования начинается 3го июня с формирования команд;
- Список участников каждой команды принимается по почте aduenko@phystech.edu;
- Каждый участник победившей команды получает 100 баллов, проигравшей – от 0 до 80 в зависимости от качества предложенных решений;
- Баллы за соревнование считаются в качестве составной части письменного экзамена;
- В каждой команде от 2 до 4 человек, общее число команд – 2;
- Очный тур, презентация алгоритмов и анализ результатов проводится во время лекции 8го июня или в другое время по согласованию;
- Данные для анализа выдаются после получения списка команд.

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 ($S_k = S_{k-1} + 2D_k$), проигравшая – теряет все конфеты ($S_k = S_{k-1} - D_k$);
- 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). Кроме того, требуется предоставить алгоритм выбора D_k, T_k для временных рядов из контрольной выборки (количество то же, что и на обучении).

Замечание 2: Алгоритм должен быть выполнен на питоне в виде функции с заданным интерфейсом и предобучен по данным из обучающей выборки. Интерфейс будет сообщен дополнительно при выдаче выборок данных.

Замечание 3: Время работы алгоритма на всей контрольной выборке (равнозначна обучающей) - не более 30 минут на среднем ноутбуке.