

# Соревнование по курсу «Байесовский выбор моделей»

## Общая информация

- Заочная часть соревнования начинается 18го ноября с формирования команд и выбора ведущего;
- Заявки на роль ведущего принимаются до 21.00 по Москве 18го ноября;
- Ведущий будет выбран до 8.00 19го ноября, после чего принимаются списки команд;
- Список участников каждой команды принимается по почте aduenko1@gmail.com;
- Каждый участник победившей команды получает 200 баллов; баллы каждой следующей команды составляют половину от баллов предыдущей (100, 50, 25 и т.д.)
- В каждой команде от 2 до 4 человек, общее число команд – не более 4;
- Очный тур проводится ведущим во время лекции 27го ноября и его результаты сообщаются лектору по электронной почте;
- Каждый участник каждой из команд может написать отзыв о работе ведущего на aduenko1@gmail.com (авторы отзывов не будут сообщены ведущему, только агрегированная статистика);
- За проведение очного тура ведущий получает 80 баллов + 10 · (# положительных отзывов - # отрицательных отзывов);
- Данные для предварительного анализа выдаются после получения списка команды;
- Данные состоят из 4х выборок: две выборки по прогнозу временного ряда, две выборки на бинарную классификацию.

## Описание процедуры соревнования

**Прогноз временного ряда (две выборки).** Имеется одномерный временной ряд  $x_1, \dots, x_N$ , сгенерированный по некоторому правилу.

**Предобучение:** для предобучения выдается первая часть временного ряда  $x_1, \dots, x_n$ , возможно содержащая пропуски (закодированы с помощью NAN).

**Описание игры:**

- Перед началом игры у каждой команды есть  $S_0 = 1000000$  конфет;
- Игра состоит из  $N - n$  шагов;
- На шаге  $t = 1, \dots, N - n$  каждая команда генерирует прогноз  $\hat{x}_{n+t}$  элемента временного ряда  $x_{n+t}$ , а также свой вклад  $D_t$  от 0 до полного числа конфет  $S_{t-1}$ , которыми команда располагает до хода  $t$ ;
- Побеждает та команда, прогноз которой есть ближайший к истинному значению временного ряда в терминах  $l_1$  расстояния;
- Вклад победившей команды удваивается ( $S_t = S_{t-1} + 2D_t$ ), вторая по близости команда возвращает исходное количество конфет ( $S_t = S_{t-1}$ ), остальные – теряют все конфеты ( $S_t = S_{t-1} - D_t$ );

- Всем алгоритмам сообщается истинное значение  $x_{n+t}$  временного ряда, а также прогнозы и вклады всех команд;
- Переход на шаг  $t + 1$ .

**Задача:** построить алгоритм прогноза следующего элемента временного ряда и назначения вклада по предыдущим значениям временного ряда, а также по данным о поведении остальных команд.

**Замечание 1:** алгоритм должен быть предобучен по предоставленным данным так, чтобы время работы на одном шаге не превышало 10мс на компьютере ведущего.

**Замечание 2:** алгоритм должен быть выполнен на питоне в виде функции с заданным интерфейсом. Интерфейс будет сообщен дополнительно при выдаче выборок данных.

**Бинарная классификация (две выборки).** Имеется выборка  $(\mathbf{x}_1, y_1, t_1), \dots, (\mathbf{x}_N, y_N, t_N)$ ,  $y_i \in \pm 1$ ,  $t_i = i$ ,  $\mathbf{x}_i \in \mathbb{R}^m$  из некоторой вероятностной модели  $p(y_i, \mathbf{w}_i | \mathbf{x}_i) = p(\mathbf{w}_i)p(y_i | \mathbf{w}_i, \mathbf{x}_i)$ . Вид модели неизвестен, параметры могут медленно меняться во времени (могут не меняться).

**Предобучение:** для предобучения выдается первая часть выборки  $(\mathbf{x}_1, y_1, t_1), \dots, (\mathbf{x}_n, y_n, t_n)$ , возможно содержащая пропуски в признаковом описании объектов (закодированы с помощью NAN).

**Описание игры:**

- Перед началом игры у каждой команды есть  $S_0 = 1000000$  конфет;
- Ведущий обладает бесконечным числом конфет;
- Игра состоит из  $N - n$  шагов;
- На каждом шаге  $s$  ведущий утверждает, что объект относится к классу 1 и предлагает 10000 конфет, если это не так;
- На шаге  $s = 1, \dots, N - n$  каждая команда оценивает объект  $x_{n+s}$  и генерирует свой вклад  $D_s$  от 0 до полного числа конфет  $S_{s-1}$ , которыми команда располагает до хода  $s$ ;
- Выбирается команда с наибольшим вкладом  $D_s$  (если таковых несколько, то случайно среди них);
- Если объект  $\mathbf{x}_{n+s}$  принадлежит классу 1, команда теряет свой вклад  $D_s$  в пользу ведущего ( $S_s = S_{s-1} - D_s$ );
- Если объект  $\mathbf{x}_{n+s}$  принадлежит классу -1, команда приобретает 10000 конфет от ведущего ( $S_s = S_{s-1} + 10000$ );
- Всем алгоритмам сообщается истинное значение  $y_{n+s}$  класса объекта  $\mathbf{x}_{n+s}$ , а также вклады всех команд;
- Переход на шаг  $s + 1$ .

**Замечание 1:** алгоритм должен быть предобучен по предоставленным данным так, чтобы время работы на одном шаге не превышало 50мс на компьютере ведущего.

**Замечание 2:** алгоритм должен быть выполнен на питоне в виде функции с заданным интерфейсом. Интерфейс будет сообщен дополнительно при выдаче выборок данных.