

# Задачи к зачету по курсу «детектирование аномалий»

Иван Шанин

ivan.shanin@gmail.com

ВМК МГУ — 15 апреля 2019 г.

## Задача 1

Монета подбрасывается 100 раз. Оцените вероятность того, что монета выпала орлом более 90 раз, используя

1. неравенство Маркова
2. неравенство Чебышева
3. неравенство Чернова
4. неравенство Хёфдинга

## Задача 2

Скачайте датасет KDD Cup 1999: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> Проведите отбор количественных признаков с помощью метода главных компонент (PCA). Подпро-  
станство какой размерности необходимо, чтобы сохранить а) 80%, б) 95% и в) 99% вариативности данных.

## Задача 3

Используя тот же датасет, детектируйте аномальные записи с помощью

1. EM-алгоритма (GMM). Проведите эксперимент по выбору оптимального числа компонент. Влияет ли диагональность матрицы ковариаций на результат?
2. Одноклассового SVM. Сравните полиномиальное ядро и RBF.
3. Изолирующего леса. Влияет ли предвадительный отбор признаков на результат?

## Задача 4

Скачайте временной ряд <https://datamarket.com/data/set/22pv/quarterly-electricity-production-in-australia!ds=22pv&display=line>. Постройте прогноз значений временного ряда с помощью модели SARIMA. Требуемая глубина прогноза — 5 лет. Не забудьте привести ряд к стационарному виду и провести анализ остатков!