

# Домашнее задание №2

## Метод главных компонент

CMF 2016

Задание нужно сделать в **R** и оформить максимально подробно, с пояснениями, красивыми графиками и вот этим всем. Срок сдачи — до 23:59 18.10, то есть до конца вторника :}

Скачайте из репозитория *Kaggle* (*link*) данные — изображения рукописных цифр. Каждое изображение имеет размер 28 на 28 пикселей и представлено вектором из  $\mathbb{R}^{784}$ .

### 1 Метод главных компонент

- примените метод главных компонент к полной выборке;
- постройте график зависимости собственных чисел от номера собственного числа (отсортированных в порядке убывания);
- постройте графики кумулятивной суммы собственных чисел (то есть доли объясненной дисперсии);
- объясните, сколько по вашему нужно взять собственных векторов, чтобы, с одной стороны, объяснить как можно больше дисперсии, а с другой — как можно больше сжать размерность данных.

Первый этап в задании — применение метода главных компонент — очень важно сделать вручную, чтобы разобраться, как этот метод работает. Так что пользоваться встроенной функцией нельзя :}

## 2 Главные компоненты

Нарисуйте собственные вектора, отвечающие наибольшим собственным числам, как изображения цифр. Используйте нормировку или линейные преобразования для улучшения контраста изображений.