

# Домашнее задание №1

## Выбор регрессоров и регуляризация

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 *UCI* ([link](#)) данные о лесных пожарах. В качестве регрессоров используйте все, кроме даты и времени; в качестве прогнозируемой величины —  $\log_{10}(\text{area} + 1)$ , где  $\text{area}$  — площадь возгорания.

Будем считать, что мы хотим построить линейную регрессию с тремя самыми значимыми признаками, и для этого надо их выбрать. Сделаем это двумя способами.

### 1 Пошаговая процедура

Выберите три наиболее значимых признака с помощью пошаговой процедуры, как с добавлением, так и с удалением регрессоров. Интерпретируйте результат.

### 2 Регуляризация

Выберите три наиболее значимых признака с помощью регуляризации (вспомните, что нужно использовать, лассо или ридж). Для этого

- задайте разбиение для параметра регуляризации  $\lambda$  (лучше это сделать в логарифмическом масштабе);

- для каждого значения параметра постройте регрессию и сохраните значения коэффициентов линейной регрессии;
- постройте графики зависимости получившихся коэффициентов линейной регрессии как функции от параметра регуляризации;
- интерпретируйте результат.

В этом задании можно пользоваться встроенными функциями! Главное – сделать это правильно, а затем – правильно интерпретировать полученные результаты.