

# Linear Regression (повторение)

## Полезные ссылки

Линейная регрессия в sklearn: `sklearn.linear_model.LinearRegression`

Линейная регрессия в statsmodels: `statsmodels.regression.linear_model.OLS`

RESET тест: `statsmodels.stats.diagnostic.linear_reset`

Лассо: `sklearn.linear_model.Lasso`

Ридж: `sklearn.linear_model.Ridge`

Эластик нет: `sklearn.linear_model.ElasticNet`

Диагностические графики

## Вариант 1

1. Файл `fev.txt` содержит информацию о ряде детей, у которых измерялся ряд показателей (FEV=Forced Expiratory Volume). Как лучше описать зависимость роста от возраста с помощью полинома? Учтите, что зависимость должна быть монотонной. Используйте LASSO и Elastic Net для фильтрации множества предикторов.
2. В файле `hills.txt` содержатся результаты 35 гонок в Шотландии. Каждая гонка содержит результат победителя, высоты трассы и ее протяженность. Постройте регрессионную модель, при необходимости трансформировав факторы.
3. В файле `Kittiwake.txt` содержится информация о площади островов и количестве пар морских птиц, которые там обитают на островах в Шотландии. Постройте линейную регрессионную модель, проведите RESET тест. Попробуйте подобрать различные нелинейные зависимости.
4. В массиве `ethanol` содержится информация об эксперименте по сжиганию этанола в двигателе. Требуется предсказать количества оксида и диоксида азота в зависимости от параметров двигателя: `compression` и `equivalence`. Построить графики остатков и диагностировать модель.

## Вариант 2

1. Файл `mtcars.txt` содержит информацию об автомобилях. Используя остальные предикторы (кроме модели), предскажите mpg (miles per gallon). Используйте LASSO и Elastic Net для фильтрации множества предикторов.
2. В файле `cherry.txt` содержится информация о параметрах дерева. Постройте модель для прогноза объема на основе высоты дерева и его диаметра, преобразовав при необходимости переменные. Постарайтесь подумать над физическим смыслом модели.
3. В файле `troutPVC.txt` содержатся данные о концентрации полихлорированного дифенила в форели, обитающей в озере Кайга, а также возрасте этой форели. Постройте линейную регрессионную модель, проведите RESET тест. Попробуйте подобрать различные нелинейные зависимости.
4. В массиве `ore` содержатся данные о толщине слоя породы в зависимости от координат измерения. Предсказать вид зависимости. Построить графики остатков и диагностировать модель.