**Лабораторная работа 3**

1. **Polynomial.ipynb**:

На теории мы выяснили, что для нелинейной зависимости у от **Х,** можно использовать модель линейной регрессии. Для этого надо ввести новые признаки, введя степени признаков. Рассмотрите примеры и выполните задания 1 – 3.

1. **weights\_heights.ipynb (Лабораторная работа 3\_1):**

Изучите материал ноутбука и выполните задания 1 – 2.

1. **Линейная регрессия.ipynb (Лабораторная работа 3\_2):**

Изучите материал ноутбука и выполните задания блок 1 – блок 4.

1. **lab\_1\_examples.ipynb (Лабораторная работа 3\_3)**

Изучите материал ноутбука и выполните задания:

* Построить несколько моделей линейной регрессии с использованием метода OLS (оставляя по одному из двух зависимых признаков; проведя нормализацию признаков, значения которых большие, и Y)
* В методах Ridge и Lasso «поиграть» параметром alpha и сравнить полученные при этом коэффициенты.

1. Набор данных house\_price.csv содержит характеристики домов и цену продажи ($).

**Задача:** построить модель, которая по числовым характеристикам жилья предскажет его цену.

Для построения модели используйте LassoCV ([https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html) [learn.org/stable/modules/generated/sklearn.linear\_model.LassoCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html)).

В LassoCV параметр alpha (λ) подбирается автоматически с помощью перекрестной проверки (кросс-валидации). Список значений для alpha можно задать явно используя параметр **alphas**. С помощью параметра **cv** устанавливают количество блоков для *k*-блочной перекрестной проверки. Атрибут **alpha\_** возвращает значение alpha, выбранное с помощью перекрестной проверки, **coef\_** возвращает коэффициенты модели, а **intercept\_** возвращает сдвиг. Метод [**score**](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html#sklearn.linear_model.LassoCV.score)возвращает коэффициент детерминации. Другие параметры, атрибуты и методы см. в [документации](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html)).

*Замечание. Параметры в классах Sklearn – это гиперпараметры, т.е. параметры, которые мы задаем при создании экземпляра класса!*

1. Загрузите данные в DataFrame, используя функцию read\_csv.
2. Сколько строк и столбцов в данных? Есть ли пропуски? Для ответа на вопросы используйте метод info().
3. Для выполнения задания в наборе данных необходимо оставить только числовые признаки.
4. Удалите столбец Id и пропущенные значения.
5. Разделите набор данных на входные данные **X** (все столбцы кроме SalePrice) и ответы **y** (столбец SalePrice).
6. Разделите данные на обучающую и тестовую выборки.
7. Обучите модель [LassoCV](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html) (установите значение гиперпараметра **cv** самостоятельно). Оцените качество полученной модели. Посмотрите на коэффициенты модели. Есть ли коэффициенты равные 0? Что это означает?

**from sklearn.linear\_model import** LassoCV

lasso = LassoCV(cv=?)

**Сделайте выводы.**

1. L2-регуляризацию: обучите модель [RidgeCV](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeCV.html).
2. Сравните полученный результат LassoCV и RidgeCV.