# Задание 1.

1. За весь пункт 4 балла. Рассмотрим первый датасет. Пусть 1 столбец данных – х, 4 – y
   1. Обучите линейную регрессию. Предскажите значения для x=1,2,3
   2. Используя кросс-валидацию и среднеквадратичную функцию потерь, оцените точность и способность к генерализации у модели
   3. Обучите полиномиальную регрессию, для кросс-валидации постройте график bias, variance в зависимости от степени полинома. Какая степень оптимальна?
   4. Проведите подбор степени полинома с помощью любого из k-fold на 70% данных. Совпадают ли результаты? Оцените точность модели на оставшихся 30%.
2. Повторите пункты a-d для трех входных переменных. 3 балла
3. Рассмотрим линейную регрессию, построенную для фиксированного набора базисных функций: , – обратная дисперсия ошибок. Рассмотрим среднеквадратичную ошибку. Используем в качестве базиса функции

Для разложения рассмотрим функции с , расположенными равномерно между -1 и 1 и – расстояние между центрами. Повторите пункты a-d для такой регрессии. Возможно, вам придется написать свой трансформер для данных (или вы найдете другой способ). 3 балла

1. \* Предположим, что независимы и одинаково распределены в соответствии с распределением Лапласа (а не в соответствии с N(0,σ2)). То есть каждый
   1. Приведите функцию потерь минимизация которой эквивалентна нахождению MLE для β в рамках вышеуказанной модели шума. Объясните эквивалентность. 1 балл
   2. Как вы думаете, почему приведенная выше модель обеспечивает более надежное соответствие данным по сравнению со стандартной моделью, предполагающей гауссово распределение шумовых членов? 1 балл
2. \* Обучите линейную регрессию с функцией ошибок из пункта 4 с помощью градиентного спуска. 5 баллов
3. \* Обучите модель линейной регрессии с любыми ухищрениями (кроме читования) на наборе данных, посвященном определению стоимости продажи здания. Можно использовать внешние источники для дополнения датасета. Необходимо предоставить код решения в виде ноутбука. Данные из внешних источников можно прикрепить файлами. Также нужно прикрепить файл с предсказанным столбцом Y. Это решение можно делать по двое. От 1 до 4 баллов за попадание в топ 100, 75, 50, 25% и 1-2 балла за код.