

## Задание 2. Линейная регрессия

Соберите с сайта какого-нибудь интернет магазина данные по стоимости жестких дисков. Постройте модель линейной регрессии, которая бы прогнозировала стоимость диска в зависимости от его параметров.

В качестве независимых переменных попробуйте использовать следующие параметры (по отдельности и группами):

1. объем диска;
2. скорость вращения;
3. объем кэш-памяти;
4. оптимизация для RAID массивов;
5. производитель.

Постройте графики зависимости стоимости от параметров.

Охарактеризуйте качество моделей, полученных при использовании одиночных параметров и их групп. Какой вариант обеспечит самое хорошее качество модели? Есть ли смысл добавить дополнительные параметры, являющиеся нелинейными трансформациями исходных?

Оформите результат работы в виде IPython Notebook.