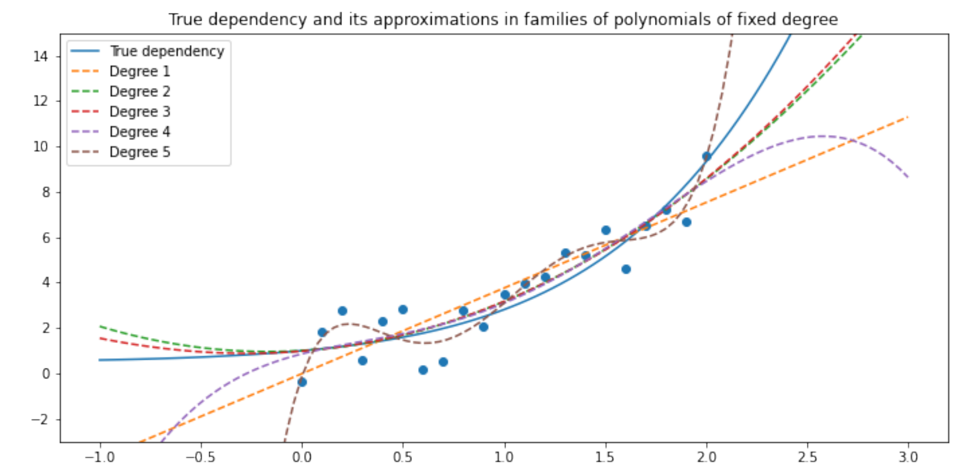
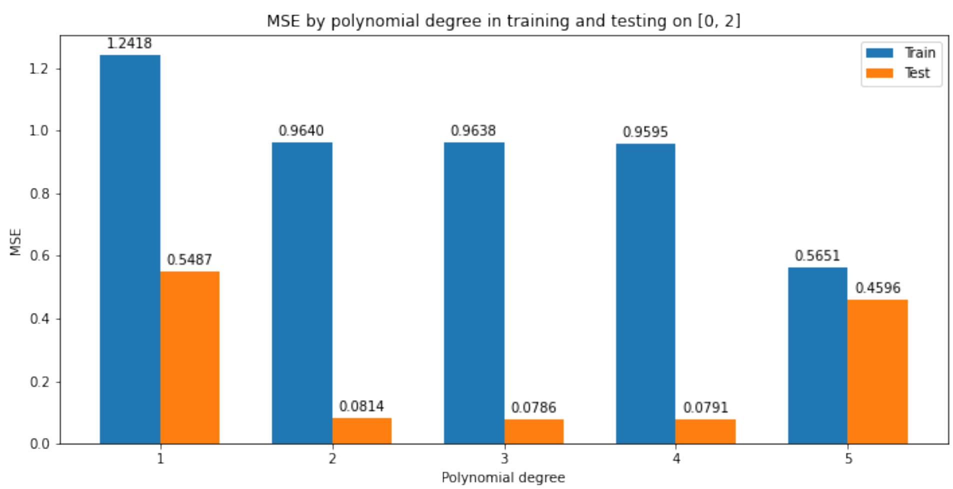
Школа анализа данных
Восстановление зависимостей
Домашнее задание №2

Кошман Дмитрий

Задача 1

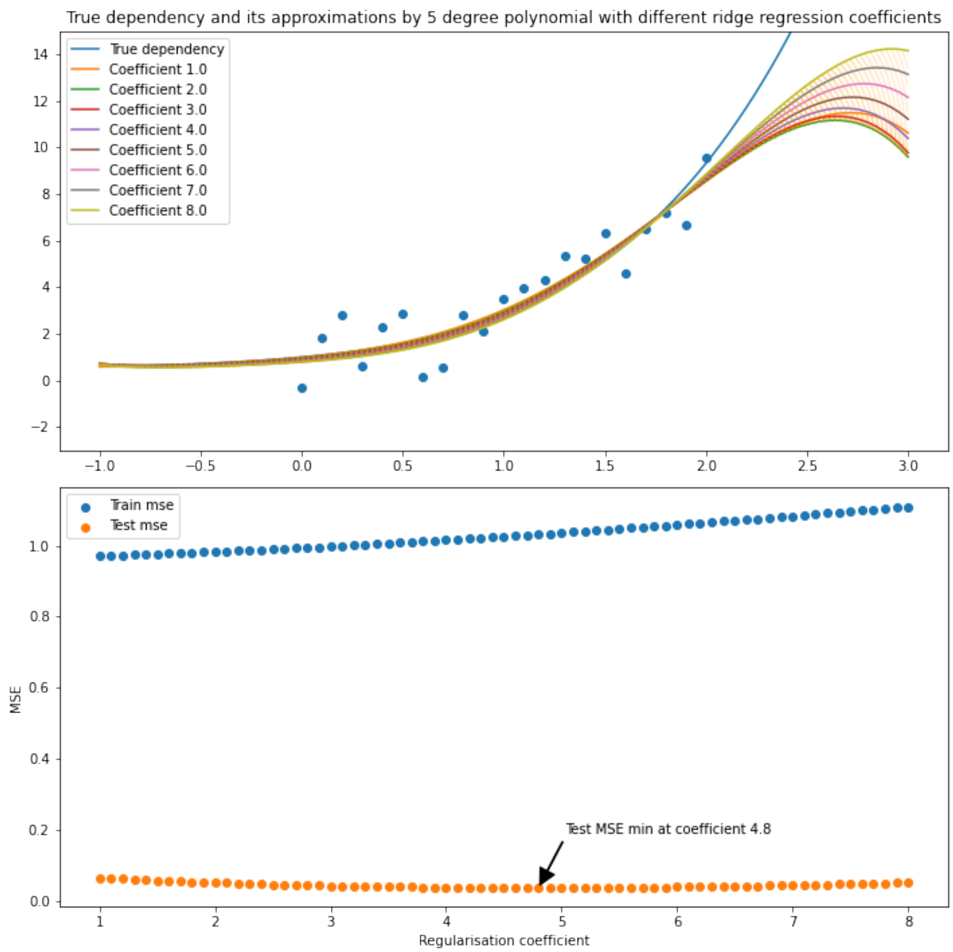
На первом графике изображены приближения зависимости полиномами разных степеней. На втором - средняя квадратичная ошибка на наблюдаемых (train) и истинних (test) ответах, сэмплированных из интервала $[0, 2]$.





Как видно, полином 5 степени лучше приближает наблюдаемые данные, что ожидаемо, поскольку он из более широкого класса функций. Но истинную зависимость лучше всего приближают полиномы 2, 3 и 4 степени, тогда как полином 1 степени не улавливает закономерностей в данных, а 5 степени подстраивается под случайный шум.

Задача 2



Наиболее точная оценка достигается при параметре регуляризации, равным 4.8.

Задача 3

Получим оценку параметра $\hat{\lambda}$ с помощью метода максимального правдоподобия:

$$\hat{\lambda} = \arg\max_{\lambda} \prod_{i=1}^{52} p_{n_i} = \arg\max_{\lambda} \sum_{i=1}^{52} \ln \frac{\lambda^{n_i}}{n_i!} e^{-\lambda} =$$

$$= \arg\max_{\lambda} \sum_{i=1}^{52} (n_i \ln \lambda - \ln n_i! - \lambda) = \arg\max_{\lambda} (-52\lambda + \ln \lambda \sum_{i=1}^{52} n_i) =$$

$$= \frac{\sum_{i=1}^{52} n_i}{52}$$

Задача 4

Получим оценку параметров $\hat{a}, \hat{M}$ с помощью метода максимального правдоподобия:

$$(\hat{a}, \hat{M}) = \arg\max_{a,M} \prod_{i=1}^{n} \frac{1}{a} [M - a/2 \leq x_i \leq M + a/2] =$$

$$= \arg\max_{a,M} \frac{1}{a^n} [M - a/2 \le x_{(1)}; x_{(n)} \le M + a/2] =$$

$$= \left(x_{(n)} - x_{(1)}, \frac{x_{(n)} + x_{(1)}}{2} \right)$$