

Практическая работа №4. Регрессионный анализ

Данные:

В работе используются наборы по прогнозированию стоимости автомобиля*:

<https://www.kaggle.com/adityadesai13/used-car-dataset-ford-and-mercedes>

Вариант $V = \text{ord}(C1) \% 9$, где $C1$ – первая буква фамилии на русском языке.

Вариант	Набор
0	audi.csv
1	bmv.csv
2	ford.csv
3	hyundi.csv
4	merc.csv
5	skoda.csv
6	toyota.csv
7	vauhhall.csv
8	vw.csv

*Можно предложить другой набор для анализа, который соответствует задаче регрессии и содержит количественные и качественные показатели.

Задачи:

1. Выполните описательный анализ (распределение показателей, корреляции с целевым показателем, пропуски, аномалии)
2. Постройте базовые регрессионные модели:
 - a. Constant (среднее значение целевого показателя)
 - b. ConstantByGroup (среднее значение на основе одного из показателей, например, среднее значение цены по модели автомобиля)
 - c. OneParamModel (линейная регрессионная модель на базе одного выбранного показателя)
3. Постройте линейные регрессионные модели для прогноза целевого показателя
 - a. Без проведения нормализации \ с нормализацией
 - b. Без регуляризации \ с регуляризацией
 - c. Только на количественных показателях \ на количественных и качественных показателях
4. Применение других регрессионных моделей
 - a. Полиномиальная регрессионная модель (с регуляризацией и без регуляризации)
 - b. Модель на основе дерева решений
 - c. Модель на основе случайного леса

Для каждой регрессионной модели вычислите оценки качества. Сделайте выводы о целесообразности применения нормализации, регуляризации моделей. Сопоставьте оценки важности показателей для разных моделей.