

Санкт-Петербургский политехнический университет
Высшая школа прикладной математики и вычислительной физики,
ФизМех

Направление подготовки
“01.03.02_01 Математическое моделирование и искусственный
интеллект”

Тема: “Прогнозирование расхода калорий”
Дисциплина: “Обработка и интерпретация сигналов”

Выполнил студент: Чуев В. Ю. (гр. 5030102/00101)

Преподаватель: Кацман В. И.

Санкт-Петербург
2024

Оглавление

1. Постановка задачи	3
2. Описание датасета.....	4
3. Метод линейной регрессии	6
4. Метод опорных векторов.....	6
5. Выводы	7

1. Постановка задачи

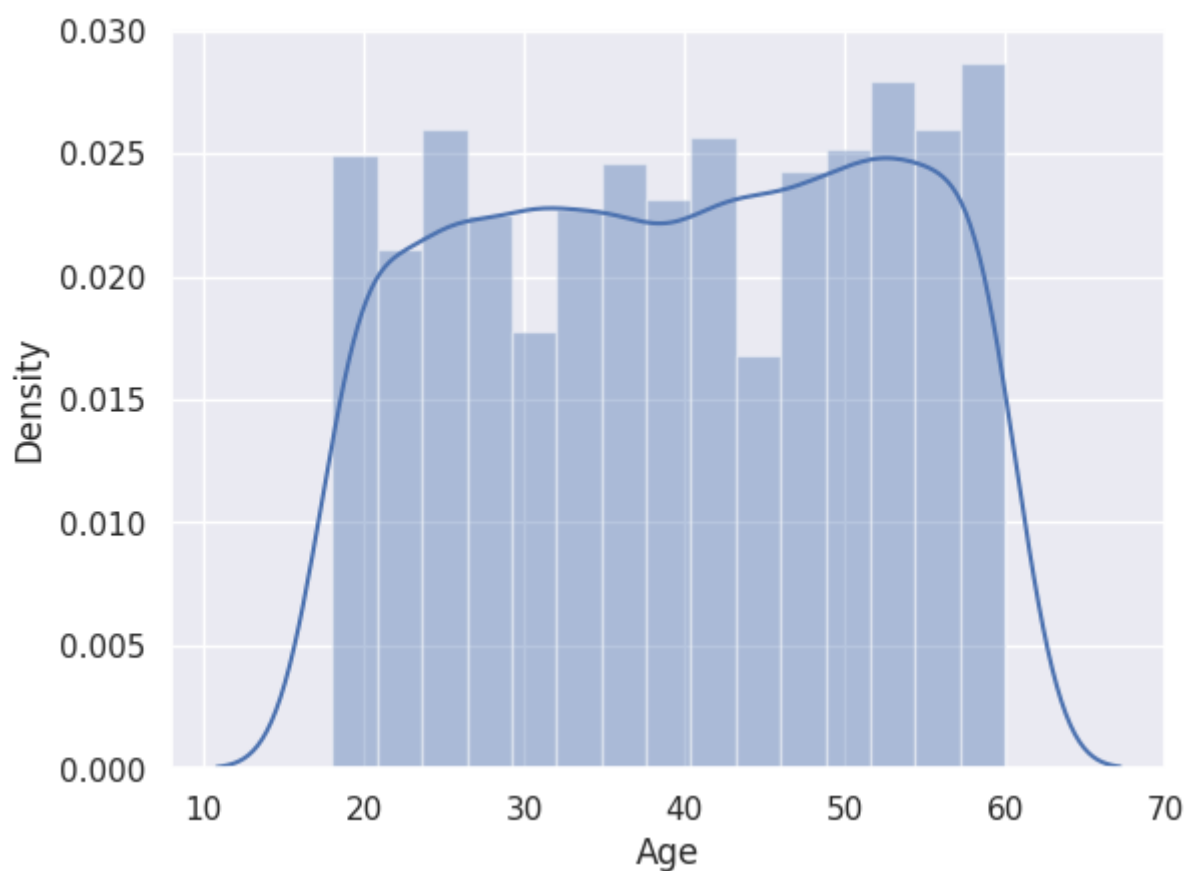
Обучить модель, которая будет получать на вход параметры человека(пол, возраст, вес, пульс, индекс массы тела), номер упражнения, время и интенсивность тренировки и выдавать ожидаемый расход калорий.

2. Описание датасета

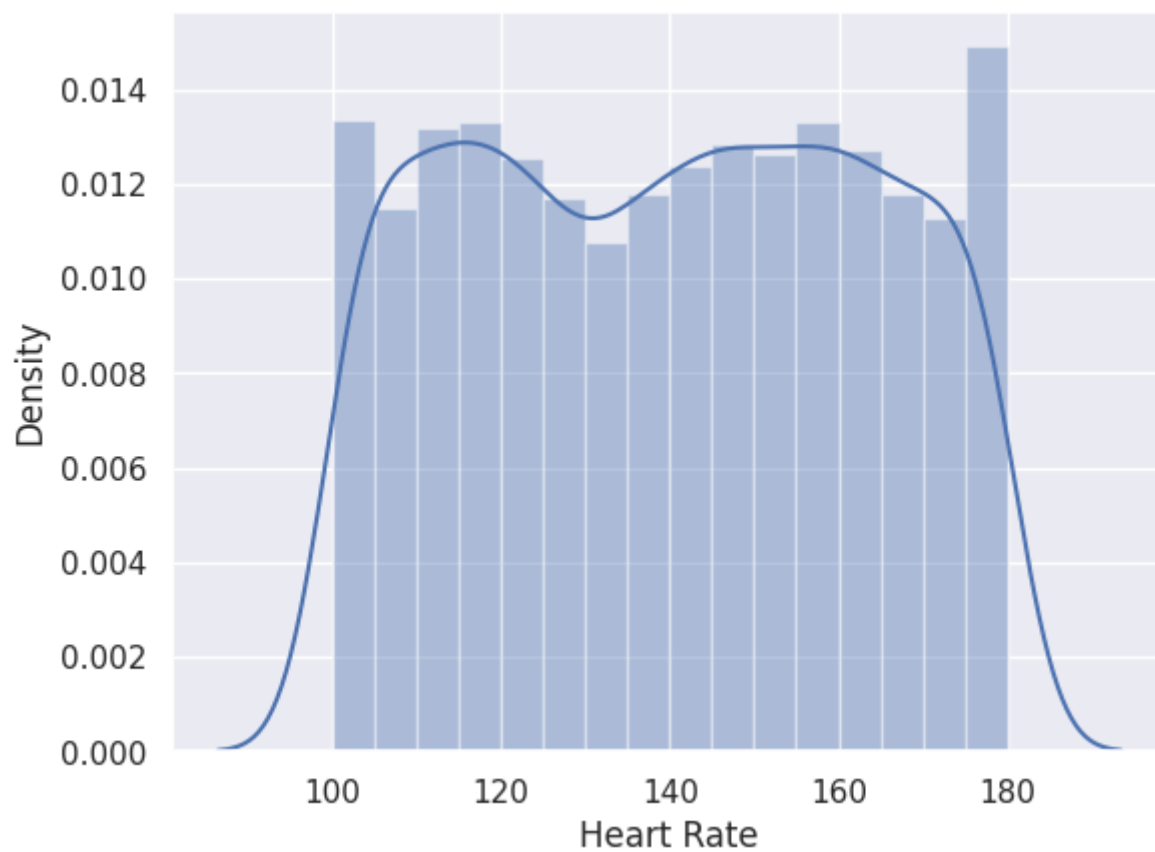
Датасет состоит из 3864 строк следующего вида:

Age	ID	Exercise	Calories Burn	Actual Weight	Age	Gender	Duration	Heart Rate	BMI	Exercise Intensity	
	0	1	Exercise 2	286.959851	96.301115	45	Male	37	170	29.426275	5
	1	2	Exercise 7	343.453036	61.104668	25	Male	43	142	21.286346	5
	2	3	Exercise 4	261.223465	71.766724	20	Male	20	148	27.899592	4
	3	4	Exercise 5	127.183858	82.984456	33	Male	39	170	33.729552	10
	4	5	Exercise 10	416.318374	85.643174	29	Female	34	118	23.286113	3

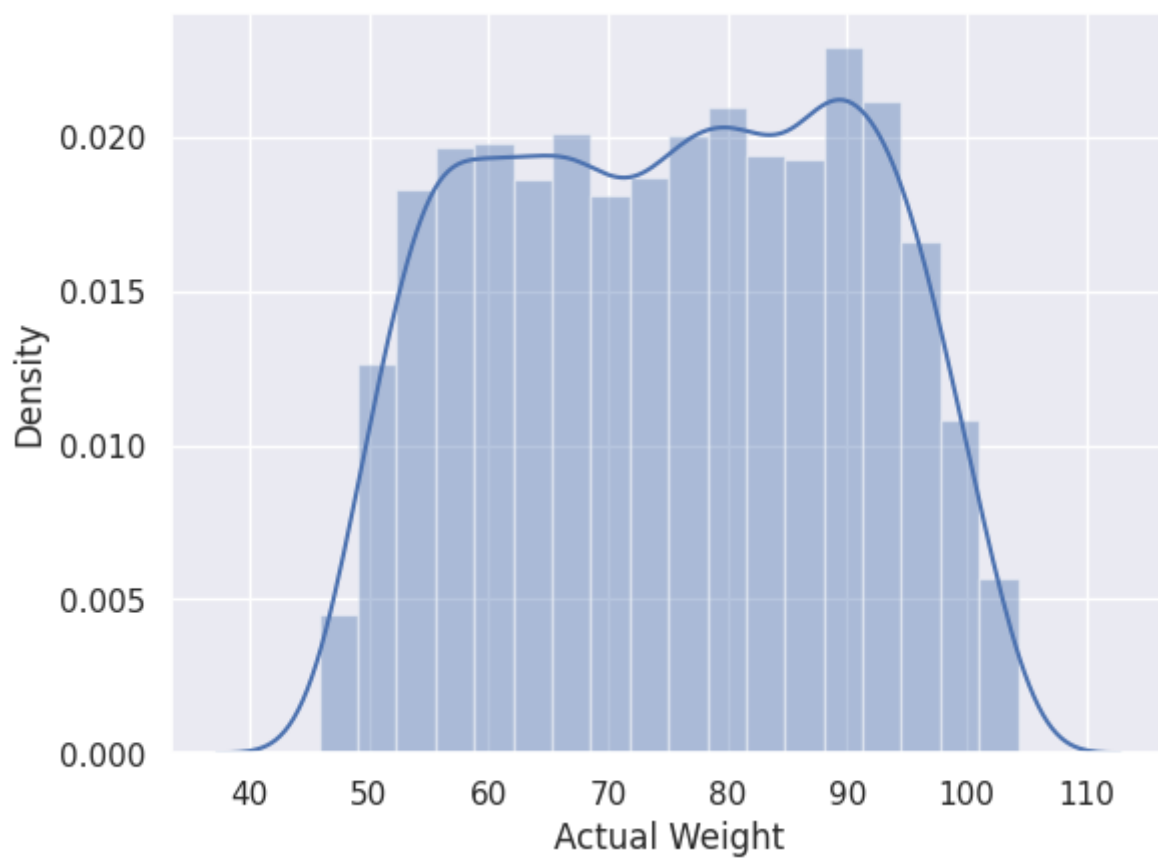
Распределение по возрасту:



Распределение по пульсу:



Распределение по весу:



3. Метод линейной регрессии

Датасет был разбит на обучающую выборку, состоящую из 3091 строк и тестовую выборку, состоящую из 773 строк и построена модель линейной регрессии. Также текстовые данные были преобразованы в числовые.

В качестве критерия качества предсказания количества калорий будем использовать среднее значение модуля разности предсказанного значения со значением из тестовой выборки(ккал), а также среднее значение модуля разности относительно значения из тестовой выборки (%)

На тестовой выборке модель показала следующий результат: средняя погрешность равна 40.1 ккал и 16.9 %. Наибольшая относительная погрешность возникает на данных, где время тренировки мало, а наибольшая абсолютная на данных, где время тренировки велико. Из этого можно сделать вывод, что такая большая погрешность возникает из-за нелинейной зависимости затраченных калорий от времени тренировки.

4. Метод опорных векторов

Построим модель методом опорных векторов, используя такое же разбиение и критерии качества. Полученные результаты: средняя погрешность равна 98.3 ккал и 43.3 %. Наибольшая погрешность возникает на тех же данных, что и в прошлом методе. Сделаем следующие итерации для улучшения этих результатов:

- 1) Масштабируем данные, чтобы среднее было равно 0, а дисперсия равна 1.
- 2) Подбираем коэффициент регуляризации $C=2000$.
- 3) Подбираем коэффициент $\gamma=0.1$ Для RBF ядра.

Изменение погрешности после сделанных итераций:

	Начальная	Масштабирование	Подбор C	Подбор gamma
Погрешность SVR				
MeanAbs(kcal)	98.3	45.4	6.14	5.75
MeanAbsPercent(%)	43.3	15.0	2.43	2.23

5. Выводы

Задача прогнозирования расхода калорий была решена с методов линейной регрессии и опорных векторов. Изначально оба метода показали невысокое качество предсказания с погрешностью 40.1 ккал/16.9% и 98.3 ккал/43.3% соответственно из-за нелинейной зависимости затраченных калорий от времени тренировки. После итераций по улучшению модели (масштабирование данных, подбор C, подбор gamma), погрешность метода опорных векторов равна 5.75 ккал/2.23%, что сопоставимо погрешности измерения расхода калорий и является хорошим результатом.