

Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Высшая школа теоретической механики и математической физики

# **ПРОГНОЗИРОВАНИЕ КРИТИЧЕСКОЙ ТЕМПЕРАТУРЫ СВЕРХПРОВОДНИКОВ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ**

---

Выполнил: А.В. Корнелюк

Группа: 5030103/00301

Научный руководитель: О.А. Троицкая

Научный консультант: А.Д. Ершов

2024

# Введение

Сверхпроводящие материалы, также известные как сверхпроводники - это материалы, которые проводят электрический ток с нулевым сопротивлением.

Сверхпроводники имеют широкое и значительное практическое применение:

- Медицинские системы МРТ;
- Сверхпроводящие катушки в Большом адронном коллайдере;
- Устройства для тонкого измерения магнитного поля.

Сверхпроводники могут изменить будущее энергетической отрасли, так как передают электроэнергию без сопротивления, а значит, без потерь.

# Актуальность

- Сверхпроводник проявляет свои ценные свойства только при температурах ниже критической температуры сверхпроводимости  $T_c$ ;
- Научная модель и теория, предсказывающие значение  $T_c$ , по сей день остаются открытыми проблемами;
- Эмпирические правила прогнозирования значения  $T_c$  не всегда работают;
- Необходим подход, основанный на **большом количестве данных**.

# Цель и задачи работы

**Цель** – создание модели машинного обучения, позволяющей эффективно прогнозировать значение критической температуры сверхпроводимости для материала, основываясь на его химических и физических свойствах.

## **Задачи:**

- Поиск и подготовка данных, создание признаков;
- Определение подхода для реализации модели;
- Построение различных моделей машинного обучения;
- Анализ результатов различных моделей.

# Поиск данных

## Wolfram and Research - Element Data

Свойство	Ед. измерения	Описание
Атомная масса	а.е.м	Суммарная масса покоя протонов и нейтронов
Первая энергия ионизации	кДж/моль	Энергия для отрыва валентного электрона от атома
Атомный радиус	пикометр ( $10^{-12}$ м)	Радиус атома
Плотность	кг/м <sup>3</sup>	Плотность при нормальных условиях
Сродство к электрону	кДж/моль	Энергия для присоединения электрона к нейтр. атому
Теплоемкость плавления	кДж/моль	Энергия для перехода из твердого состояния в жидкое
Теплопроводность	Вт/(м · К)	Коэффициент теплопроводности
Валентность	-	Число химических связей, образуемых элементом

# Подготовка и очистка данных

## База данных о сверхпроводниках NIMS

[Home](#) | [Oxide & Metallic Menu](#) | [Organic Menu](#) | [Help](#)

### OXIDE & METALLIC Search System

Select Input search element

Element :   SUBST  MATTER

Select Structure

Quick search : OXIDE   Metallic

Select from all :

Select Property

Property :

Before :

Year : After :

from :  to

Detail :

Search

Reset

**Итого: 21 263 строки данных**

# Создание признаков

Признак	Формула	Пример значения
Среднее арифметическое	$= \mu = (t_1 + t_2)/2$	35.5
Взвешенное среднее	$= \nu = p_1 t_1 + p_2 t_2$	44.43
Среднее геометрическое	$(t_1 t_2)^{1/2}$	33.23
Взвешенное среднее геометрическое	$(t_1)^{p_1} (t_2)^{p_2}$	43.21
Энтропия	$-w_1 \ln(w_1) - w_2 \ln(w_2)$	0.63
Взвешенная энтропия	$-A \ln(A) - B \ln(B)$	0.26
Разность	$t_1 - t_2 \quad (t_1 > t_2)$	25
Взвешенная разность	$p_1 t_1 - p_2 t_2$	37.86
Стандартное отклонение	$[(1/2)(t_1 - \mu)^2 + (t_2 - \mu)^2]^{1/2}$	12.5
Взвешенное стандартное отклонение	$[p_1(t_1 - \mu)^2 + p_2(t_2 - \mu)^2]^{1/2}$	8.75

8 свойств x 10 статистик = 80 признаков + признак количества элементов. **Итого: 81 столбец**

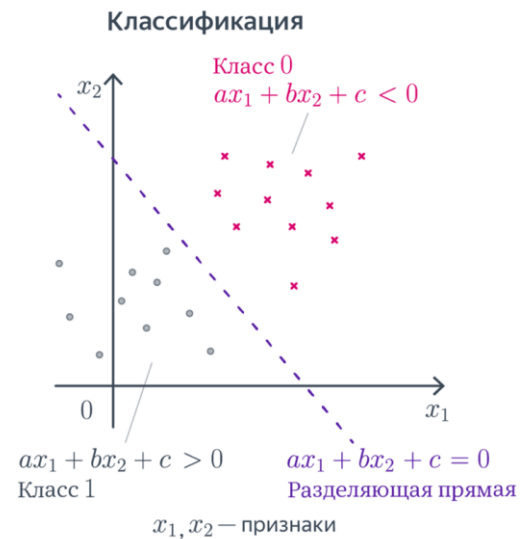
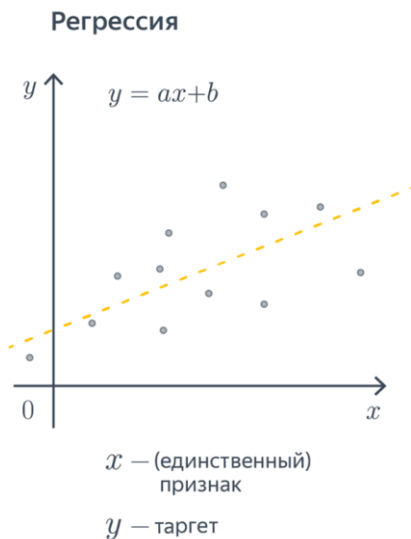
# Определение подхода для реализации модели

Типы задач машинного обучения:

- Обучение с учителем
- Обучение без учителя
- Обучение с подкреплением

Задачи обучения с учителем:

- Регрессия
- Классификация





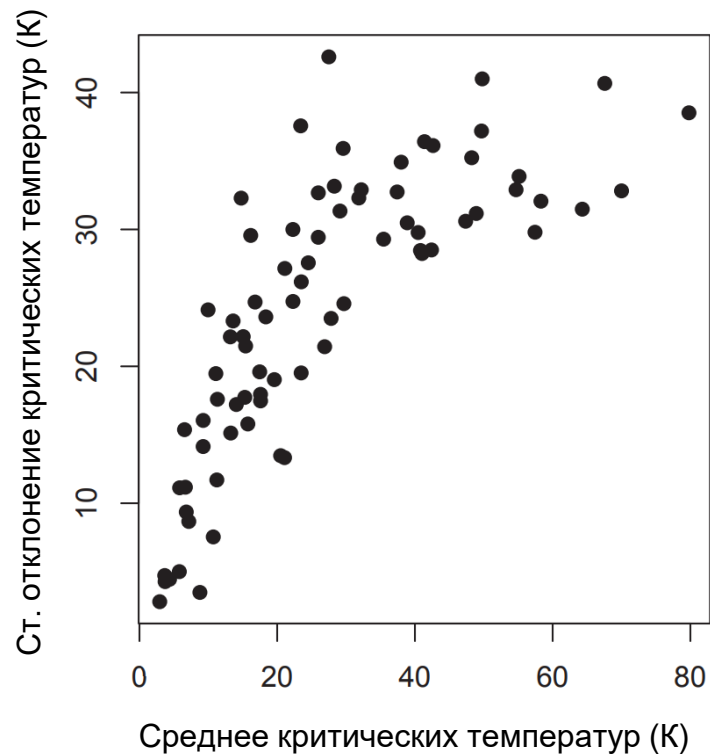
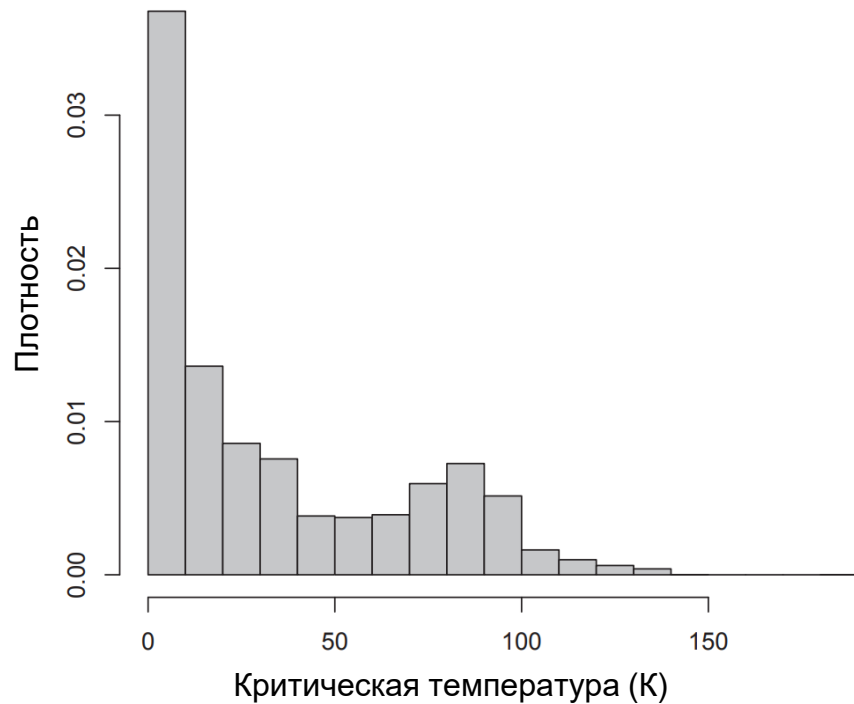
# Постановка задачи регрессии

- Имеется датасет - матрица  $X \in R^{n \times m}$  ( $n$  строк объектов,  $m$  столбцов признаков), а также вектор  $Y \in R^n$  значений целевой переменной
- Необходимо построить отображение вида  $f(X_i) \rightarrow Y_i$ , описывающее зависимость между признаками объектов и значением целевой переменной
- Для оценки качества приближения вводятся функции потерь, например:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

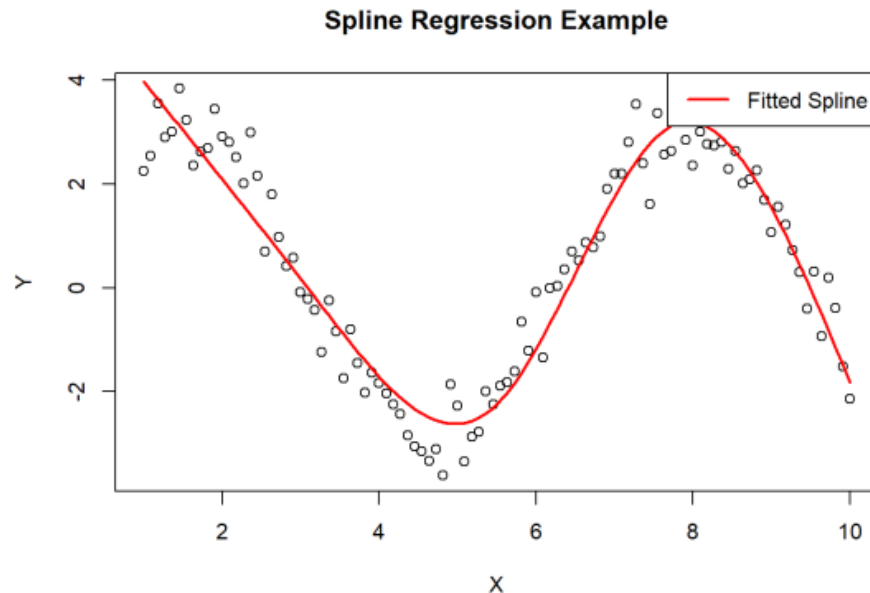
где  $y_i$  - истинные значения целевой переменной,  $\hat{y}_i$  - прогнозные значения модели, а  $N$  - число объектов в выборке, на которой обучается модель.

# Анализ данных



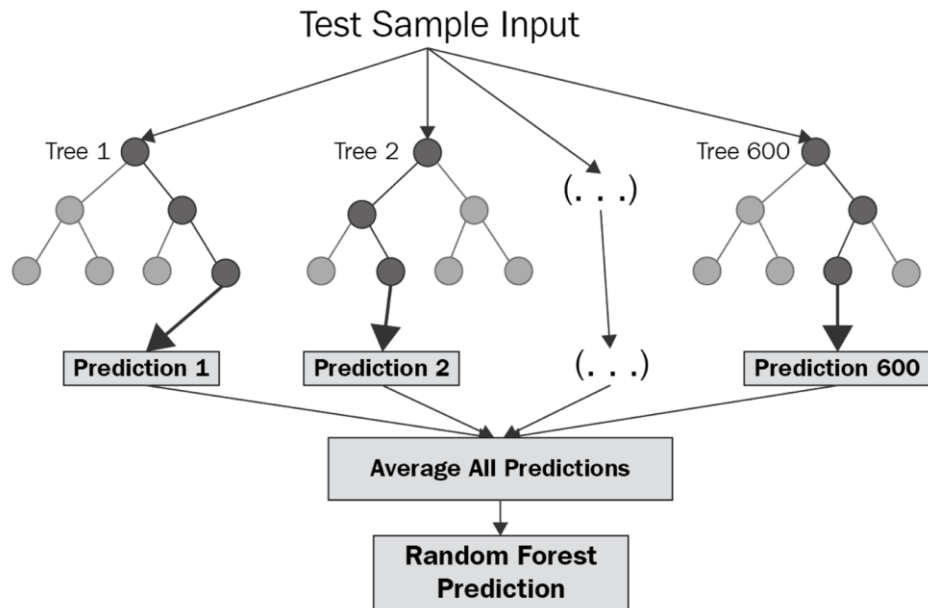
# Модель сплайн-регрессии с регуляризацией

- Суть сплайнов в следующем:  
подбор отдельных полиномов  
к различным областям значений  
целевой переменной;
- Сплайны добавляют нелинейность  
в линейную регрессию;
- Регуляризация ограничивает рост  
коэффициентов признаков.



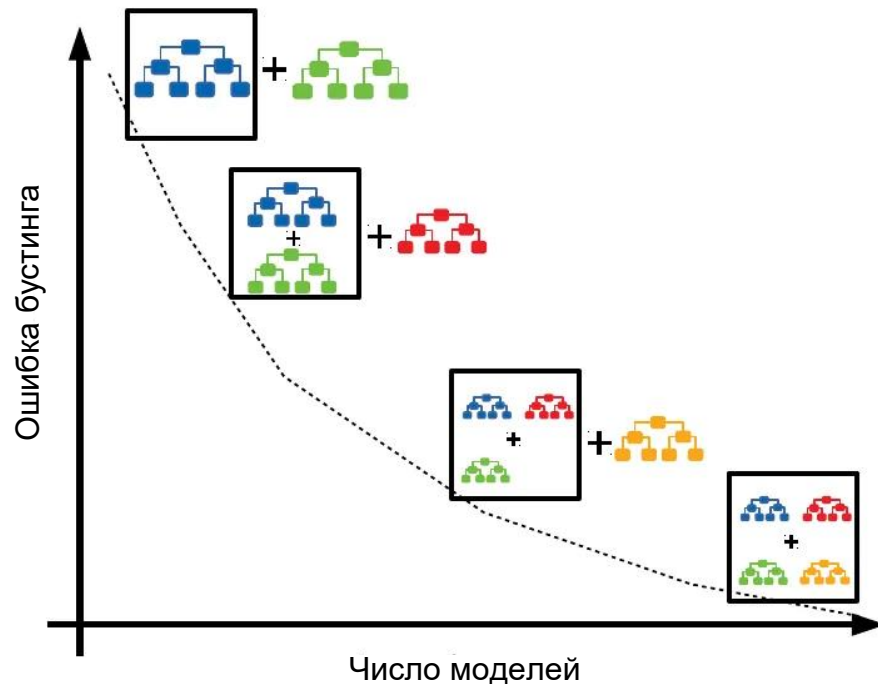
# Модель «случайный лес»

- Используется не одна, а сразу множество моделей;
- Базовая модель - дерево решений;
- Каждая модель обучается на подвыборке признаков и объектов;
- Результаты моделей усредняются.



# Модель градиентного бустинга

- Используется не одна, а сразу множество моделей;
- Базовая модель - дерево решений
- Каждая модель обучается на ошибках предыдущей модели;
- Результат бустинга - суммарный результат всех моделей.



# Результаты моделей

- Проведен подбор оптимальных параметров моделей;
- Результаты получены с помощью перекрестной проверки;
- Наилучшего результата по RMSE добилась модель градиентного бустинга.

Рейтинг	Модель	<i>RMSE</i>	<i>R<sup>2</sup> score</i>
1	XGBoost	<b>8.89</b>	0.89
2	«Случайный лес»	9.06	<b>0.90</b>
3	Сплайн-регрессия с регуляризацией	13.59	0.83

\* *R<sup>2</sup> score* показывает долю дисперсии в данных, объясненную моделью

# Оценка важности признаков

Характеристики, полученные на основе:

- теплопроводности
- атомного радиуса
- атомной массы
- валентности
- сродства к электрону

имеют наибольшее влияние на значение критической температуры

Признак	Значение Gain
Разность теплопроводностей	0.295
Взвешенное станд. отклон. теплопроводностей	0.084
Разность атомных радиусов	0.072
Взвешенное геом. среднее теплопроводностей	0.047
Станд. отклон. теплопроводностей	0.042
Взвешенная энтропия валентностей	0.038
Взвешенное станд. отклон. сродства к электрону	0.036
Взвешенная энтропия атомных масс	0.025
Взвешенное среднее валентностей	0.022
Взвешенное геом. среднее сродства к электрону	0.021
Взвешенная разность сродства к электрону	0.016
Взвешенное среднее теплопроводностей	0.015
Взвешенное геом. среднее валентностей	0.014
Станд. отклон атомных масс	0.013
Станд. отклон. плотностей	0.010
Взвешенная энтропия теплопроводностей	0.010
Взвешенная разность теплопроводностей	0.010
Взвешенное среднее атомных масс	0.009
Взвешенное станд. отклон. атомных масс	0.009
Среднее геометрическое плотностей	0.009

# Результаты работы

В работе достигнуты лучшие метрики качества, чем в аналогичных работах по прогнозированию критической температуры с помощью машинного обучения.

Было определено, что признаки на основе теплопроводности являются наиболее важными для определения критической температуры.

Данная работа может быть полезна:

- исследователям-практикам, занимающимся проведением экспериментов над сверхпроводниками с целью определения их критической температуры;
- исследователям-теоретикам, занимающимся составлением теории сверхпроводимости и вопросами поиска новых сверхпроводящих материалов.