

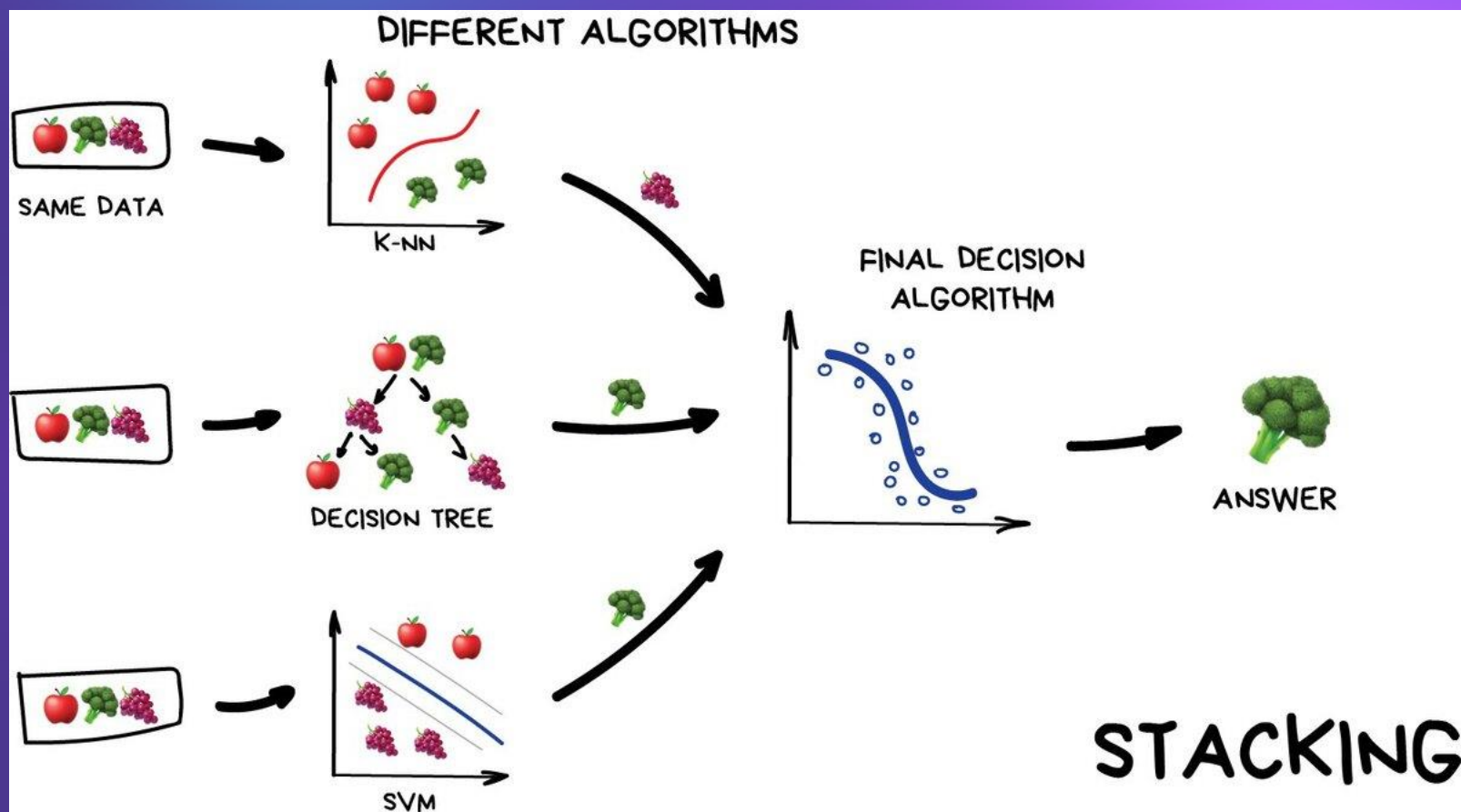


АНСАМБЛИРОВАНИЕ И РЕГУЛЯРИЗАЦИЯ

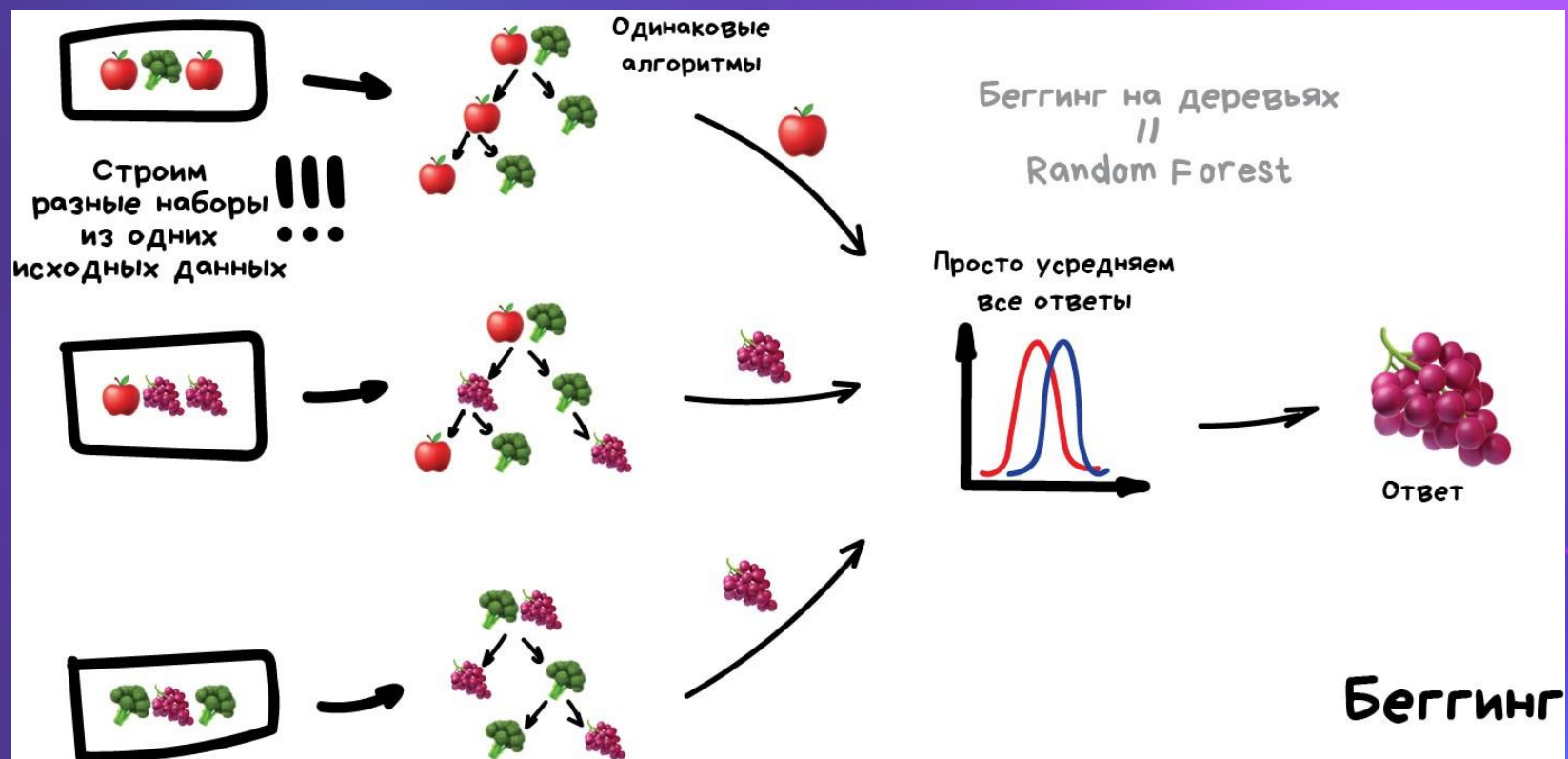
ПЛАН ВЕБИНАРА

- Повторим, что такое ансамбли и какие они бывают
- Поделаем упражнения на понимание терминов
- Узнаем, как получать от дерева информацию о том, какой признак наиболее важен
- Разберемся, как избежать переобучения в линейных моделях
- Разберем частые проблемы, возникающие, при выполнении задачи
- Разберем вопросы по теме

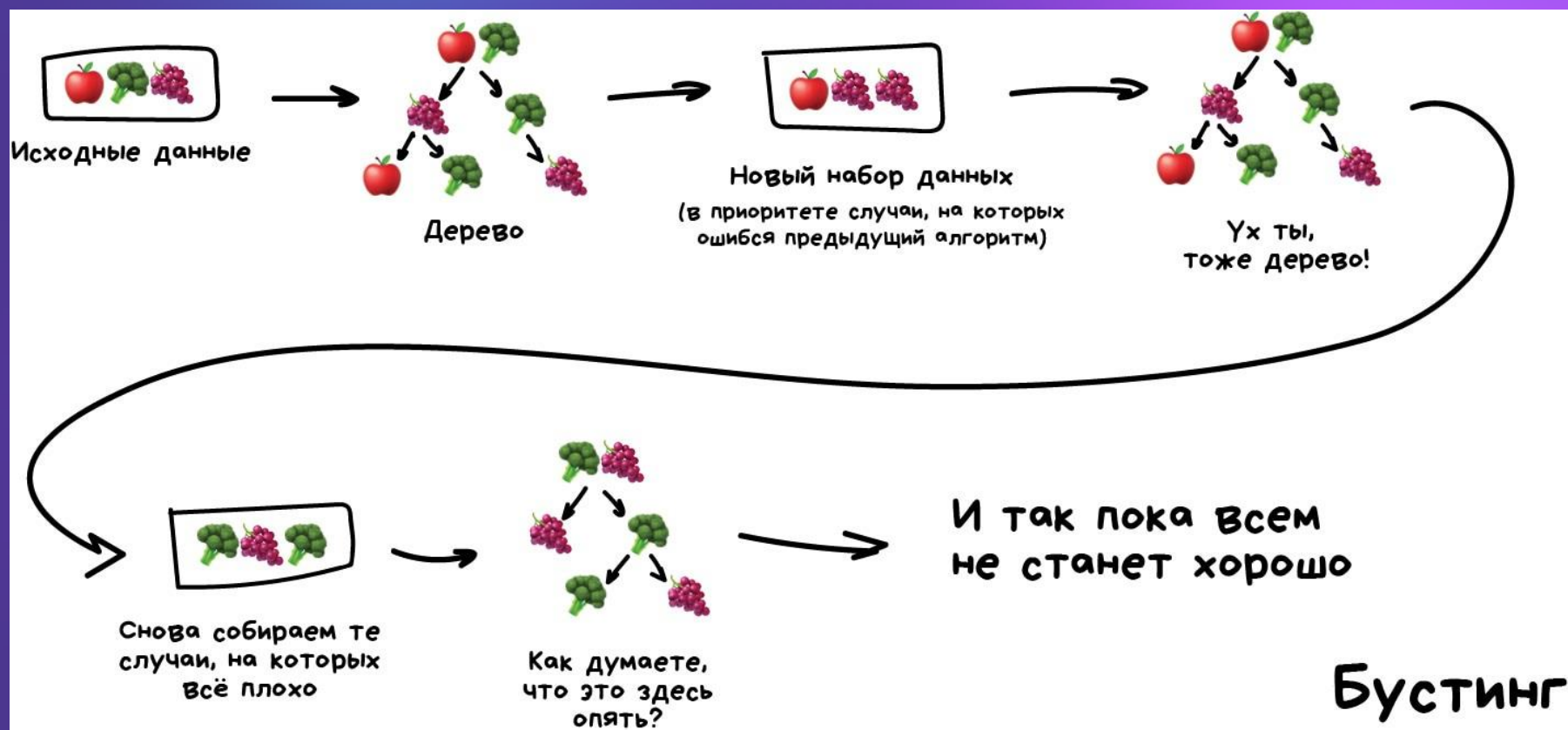
СТЕКИНГ



БЕГГИНГ



БУСТИНГ



АНСАМБЛИ НАД РЕШАЮЩИМИ ДЕРЕВЬЯМИ

- Бутстрап. Формирование из исходных данных подвыборок путем копирования.
- Бэггинг. Однородные модели обучают на разных наборах данных, а прогноз получают, усредняя ответы моделей / путем голосования.
- Случайный лес. Вариант бэггинга, у которого слабые ученики - деревья решений.
- Бустинг. Деревья обучаются последовательно, исправляя ошибки предыдущих. Прогноз получается как линейная комбинация прогнозов построенных деревьев.
- Градиентный бустинг. Вместо ошибок уже обученных деревьев пытается предсказать антиградиент функции потерь для них.

АНСАМБЛИ | УПРАЖНЕНИЕ 1

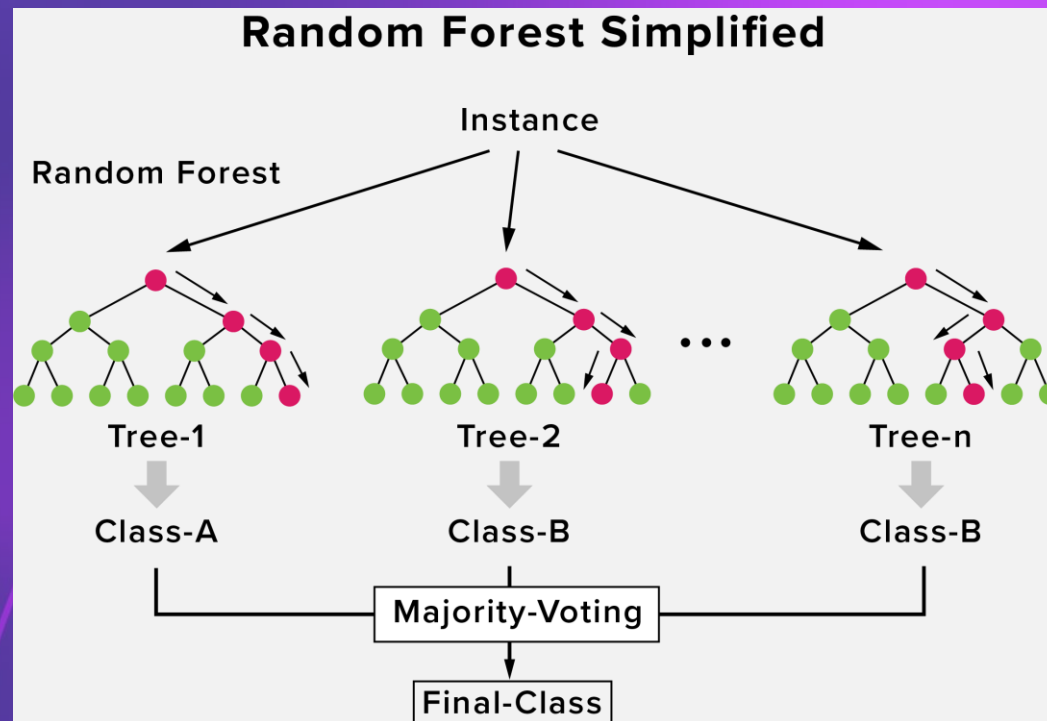
1. Каким видом ансамбля является случайный лес?

- А. Стекинг
- В. Беггинг
- С. Бустинг

АНСАМБЛИ | УПРАЖНЕНИЕ 1

1. Каким видом ансамбля является случайный лес?

- A. Стекинг
- B. Беггинг
- C. Бустинг



АНСАМБЛИ | УПРАЖНЕНИЕ 2

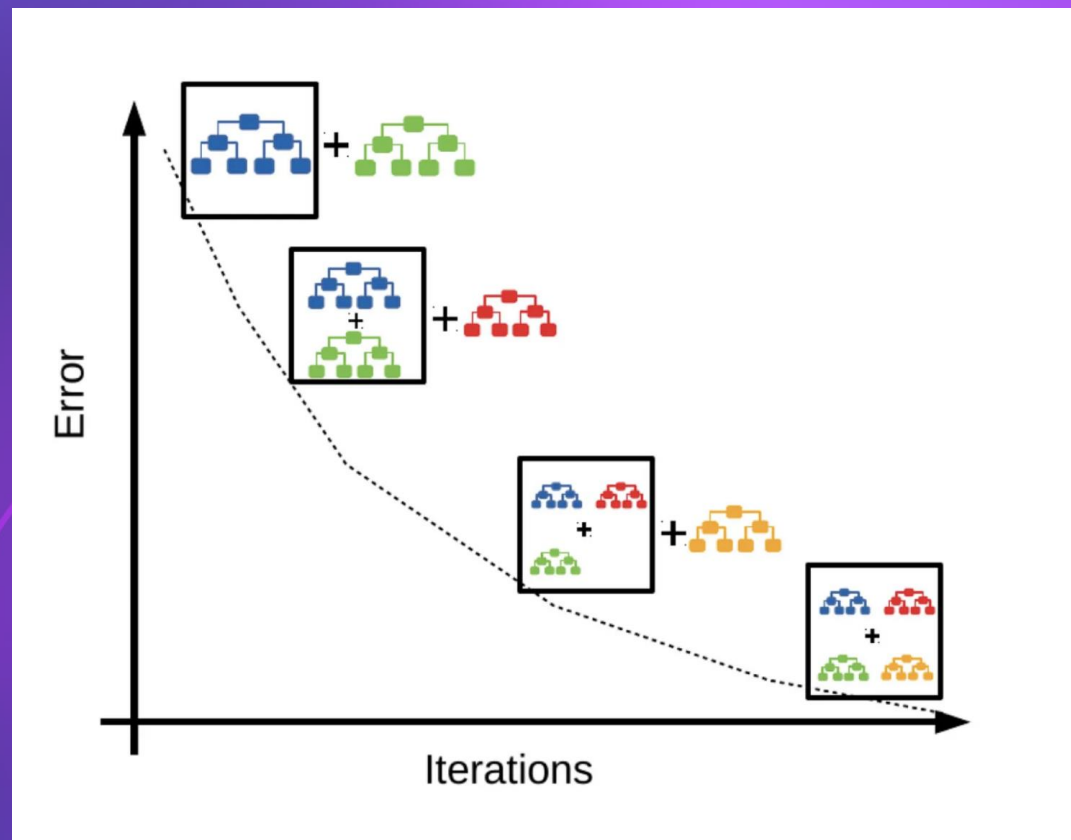
1. Что предсказывает очередное дерево в градиентном бустинге?

- A. Исходные значения целевой переменной
- B. Суммарную ошибку всех предыдущих деревьев
- C. Антиградиент общей ошибки предыдущих деревьев

АНСАМБЛИ | УПРАЖНЕНИЕ 2

1. Что предсказывает очередное дерево в градиентном бустинге?

- A. Исходные значения целевой переменной
- B. Суммарную ошибку всех предыдущих деревьев
- C. Антиградиент общей ошибки предыдущих деревьев



АНСАМБЛИ | УПРАЖНЕНИЕ 3

1. Какую задачу решает очередное дерево в градиентном бустинге?

- A. Регрессии
- B. Классификации
- C. Оптимизации

АНСАМБЛИ | УПРАЖНЕНИЕ 3

1. Какую задачу решает очередное дерево в градиентном бустинге?

- A. Регрессии
- B. Классификации
- C. Оптимизации

РЕГУЛЯРИЗАЦИЯ ЛИНЕЙНЫХ МОДЕЛЕЙ

- Позволяет избежать построения слишком сложной модели
- Получается модель с более сильными обобщающими способностями
- L1-регуляризация (Lasso) -

$$J_{Lasso}(\theta) = MSE + \alpha \cdot ||\theta||_1$$

- L2-регуляризация (Ridge) -

$$J_{Ridge}(\theta) = MSE + \alpha \cdot ||\theta||_2^2$$

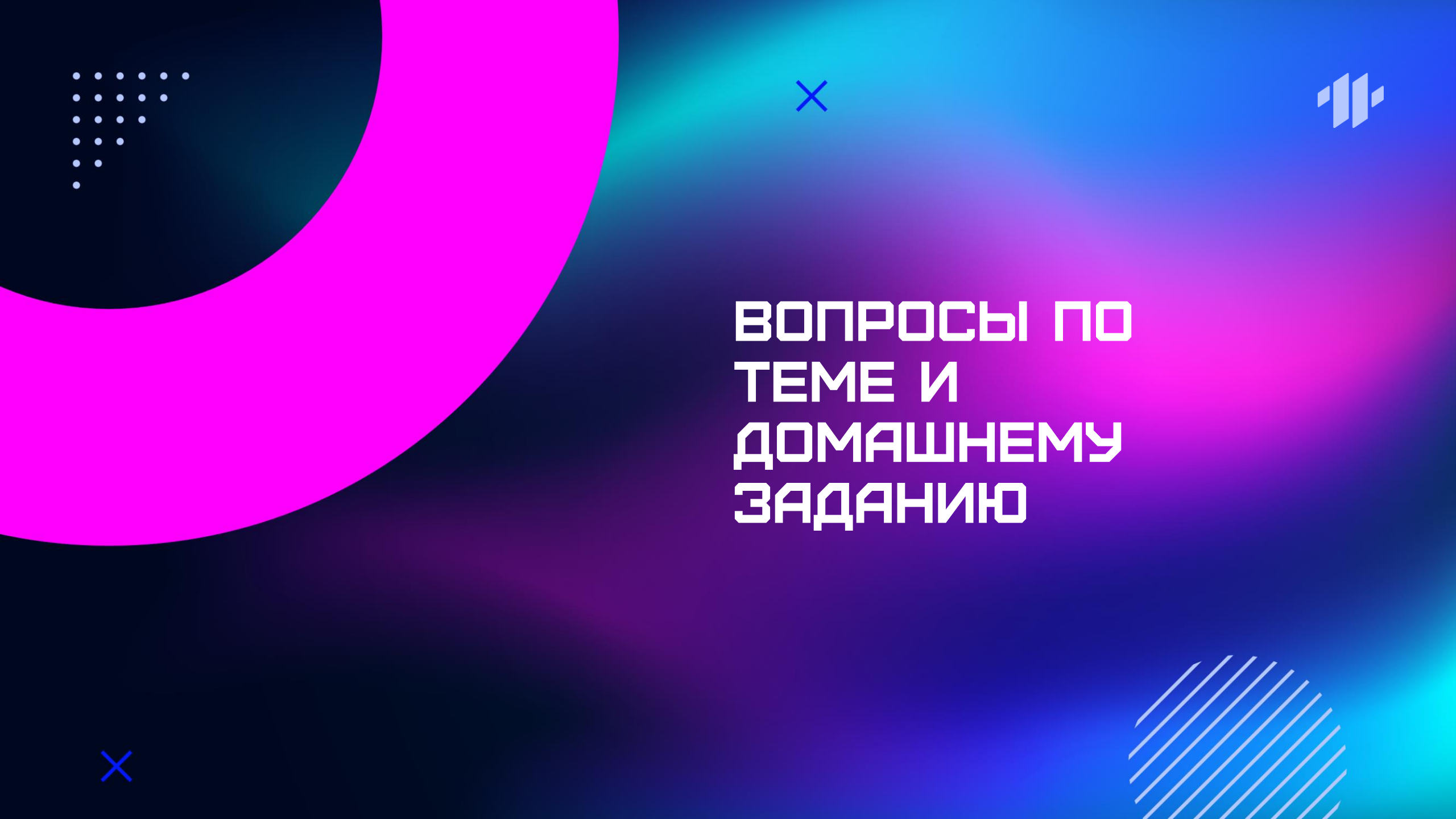
- Elastic-Net -

$$J_{ElasticNet}(\theta) = MSE + \alpha_1 \cdot ||\theta||_1 + \alpha_2 \cdot ||\theta||_2^2$$

- Все реализовано в `sklearn.linear_model`: Ridge, Lasso, ElasticNet

ЧАСТЫЕ ПРОБЛЕМЫ, ВОЗНИКАЮЩИЕ ПРИ ВЫПОЛНЕНИИ ЗАДАЧИ

- Необходимо самостоятельно сгенерировать новый целевой признак
- Функция `balanced_accuracy_score` реализуется перегружено. Можно забирать значения из матрицы ошибок, например: [так](#).
- Функция `balanced_accuracy_score` реализована, но не проверена
- Решается задача не классификации, а регрессии
- Градиентный бустинг обучается со слишком большим шагом



ВОПРОСЫ ПО
ТЕМЕ И
ДОМАШНЕМУ
ЗАДАНИЮ