What Can Transformers Learn In-Context? A Case Study of Simple Function Classes

Mark Litvinov

Few-Shot

Few-Shot Learning — это подход в машинном обучении, который позволяет модели

успешно выполнять задачи обучения, используя очень мало обучающих примеров. Это

важно в ситуациях, где доступ к большим объемам аннотированных данных ограничен

или сбор данных дорог.

- Предобучение
- Дообучение по малому числу кейсов
- Изучение способности обобщения

Мотивация

Исследование способностей трансформеров на простых алгоритмических задачах имеет несколько значительных аспектов:

- Понимание возможностей трансформеров
- Проверка гипотезы о внутриконтекстном обучении
- Расширение применения трансформеров

Какова способность трансформеров к обучению и выполнению алгоритмических задач?

Могут ли трансформеры эффективно заменить традиционные алгоритмы?

Как влияет предобучение на способности трансформеров к Few-Shot Learning в алгоритмических задачах?

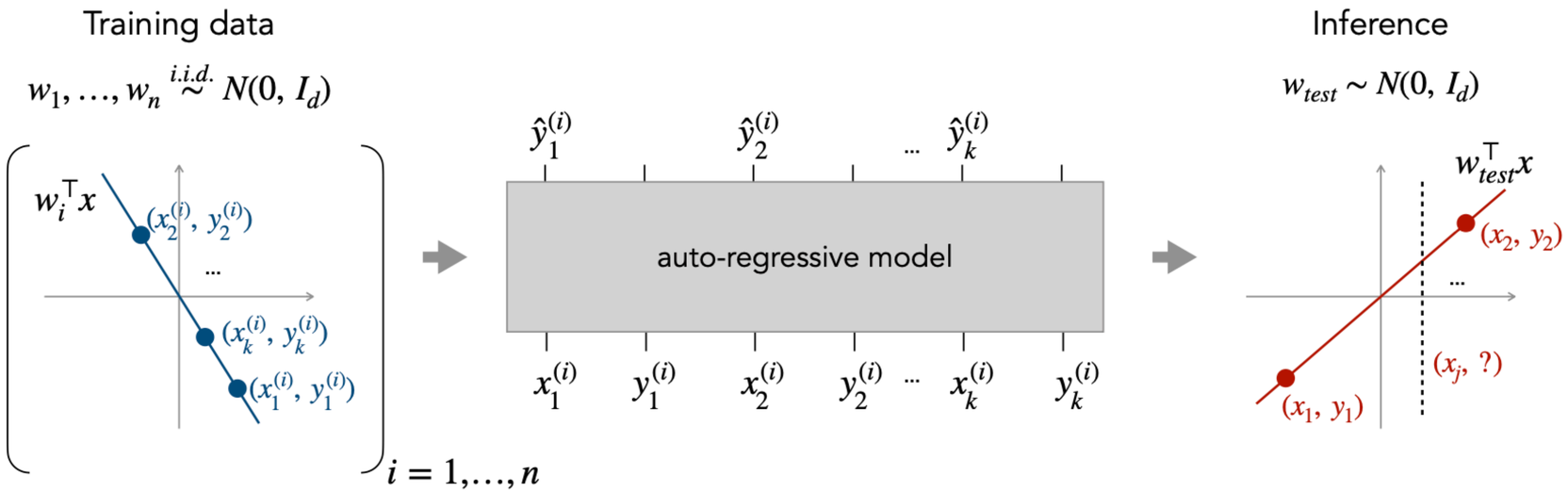
Методология исследования

Задачи для тестирования:

- Сортировка чисел
- Выполнение арифметических операций
- Поиск пути в графе
- Игры на логическом выводе

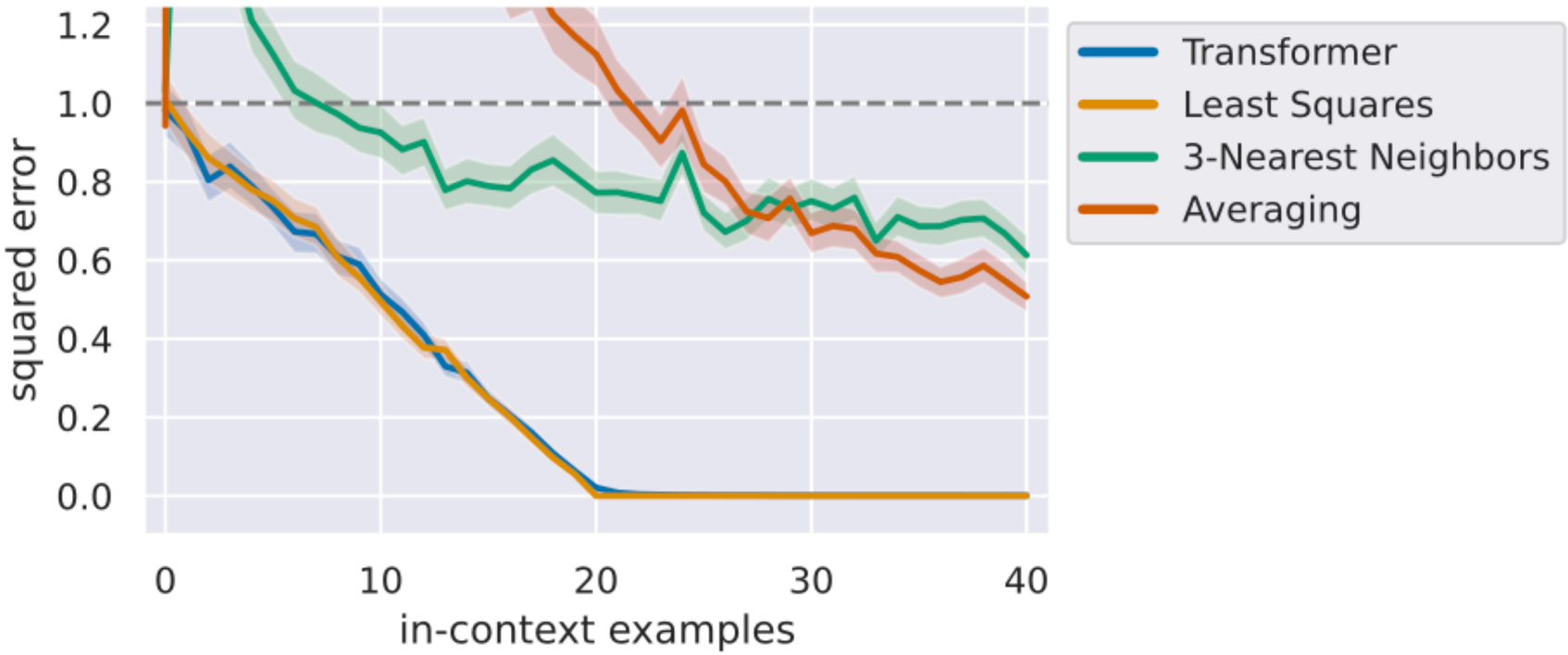
Данные: Использовались синтетически сгенерированные данные, соответствующие каждой из задач, чтобы создать стандартизированные условия для тестирования.
Процедура обучения: Трансформеры обучались на ограниченном наборе примеров, представляющих каждую задачу. Это обучение включало исключительно показ новых примеров, без повторения.
Оценка: После обучения трансформеры оценивались на основе их способности решать задачи в новых условиях, что включало изменение параметров или усложнение задач.
Эт

Обучение трансформера

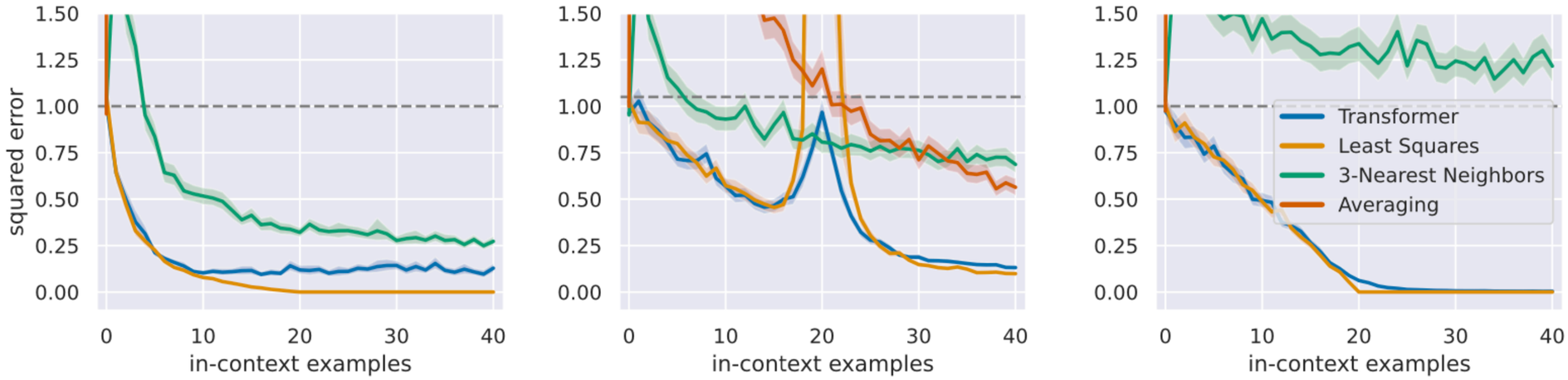


$$\min_{\theta} \mathbb{E}_P \left[\frac{1}{k+1} \sum_{i=0}^{k} \ell \left(M_\theta \left(P^i \right), f\left(x_{i+1} \right) \right) \right],$$

Результаты для линейных функций



Добавление разного шума



(a) Skewed covariance (b) Noisy linear regression (c) Different orthants

Figure 4: *In-context learning on out-of-distribution prompts.* We evaluate the trained model on prompts that deviate from those seen during training by: (a) sampling prompt inputs from a non-isotropic Gaussian, (b) adding label noise to in-context examples, (c) restricting in-context examples to a single (random) orthant. In all cases, the model error degrades gracefully and remains close to that of the least squares estimator, indicating that its in-context learning ability extrapolates beyond the training distribution.

QnA