

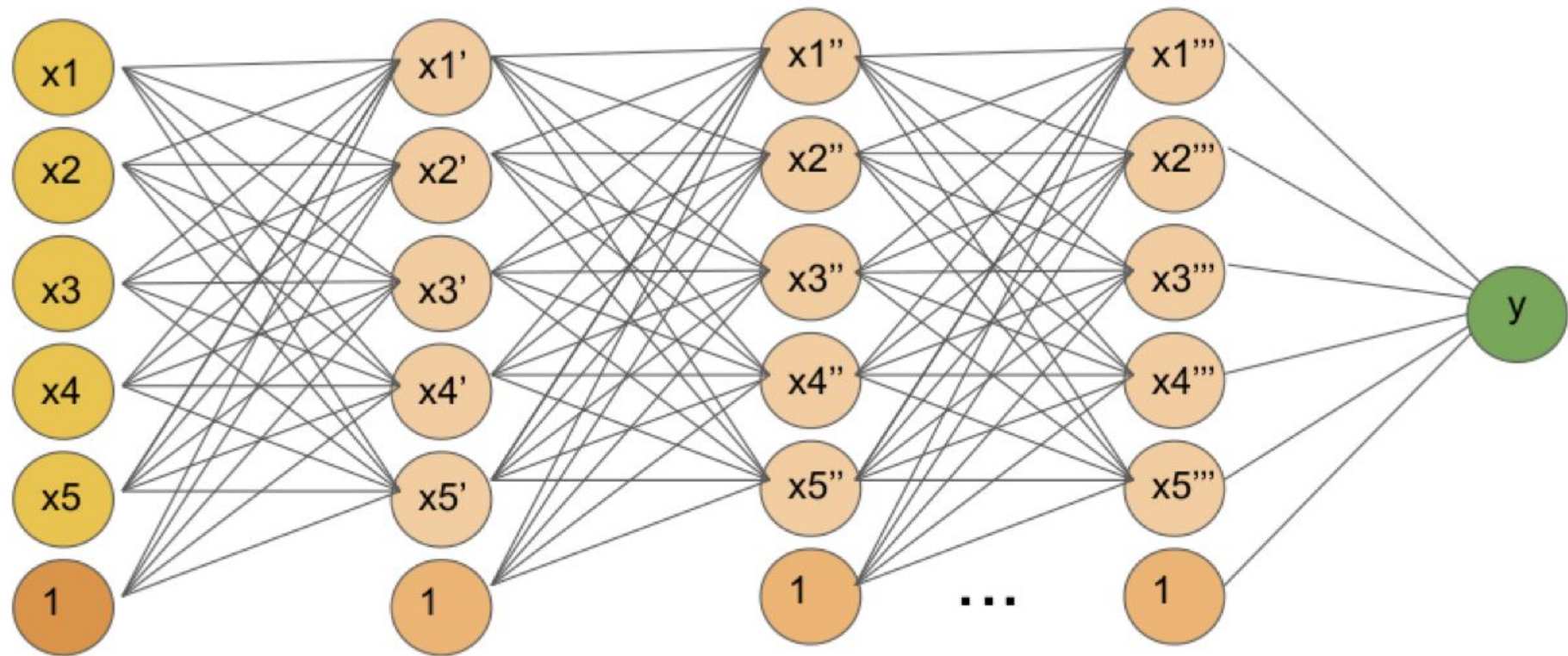
The Lottery Ticket Hypothesis

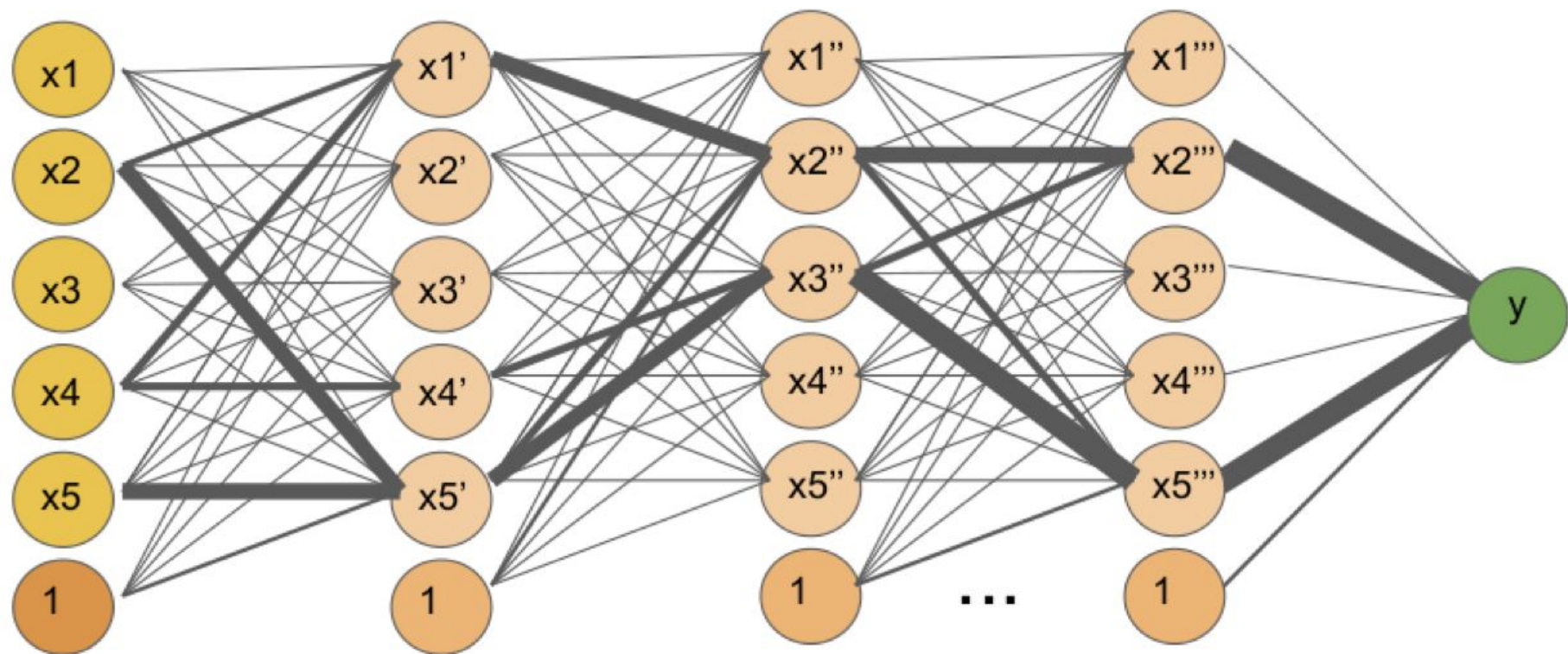
Шинкарев Елисей Сергеевич
БПМИ 211

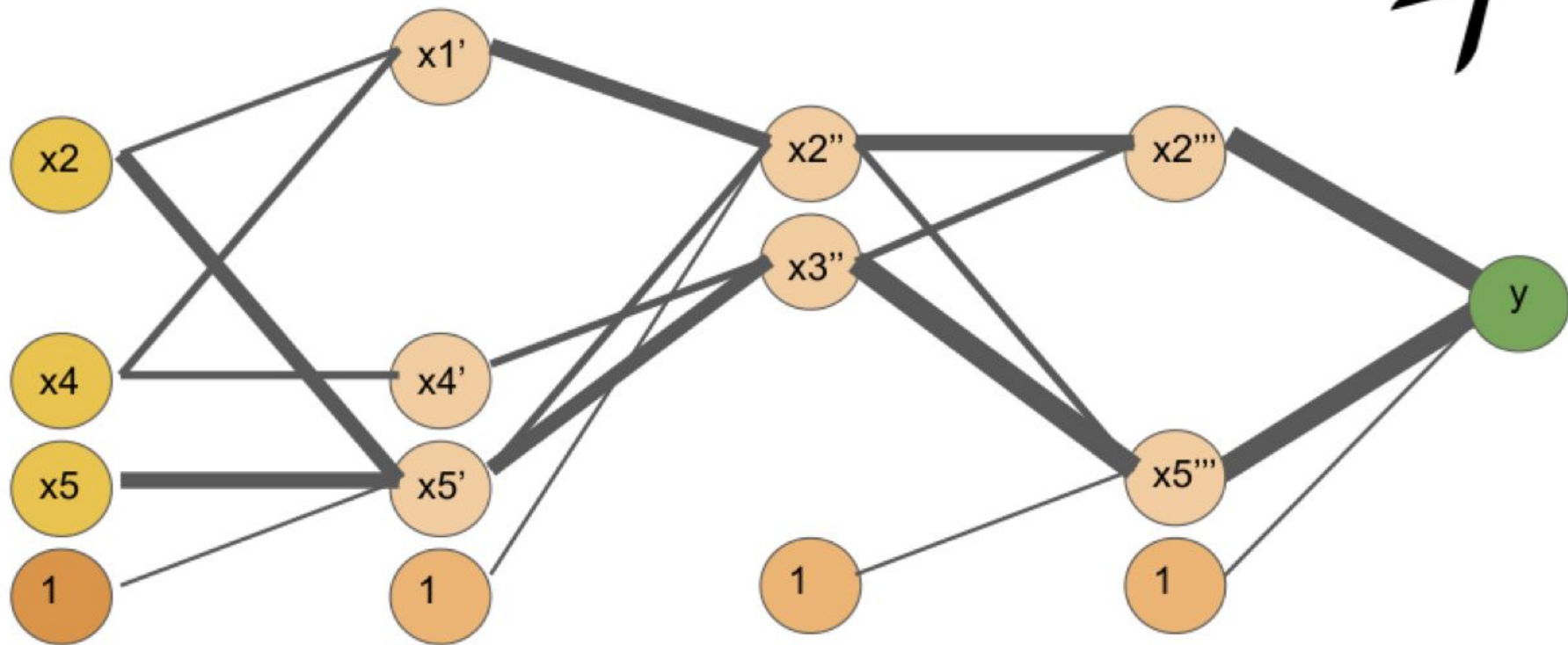
Зачем нужен прунинг и что это такое?

- Эффективность
- Память
- Скорость
- Качество

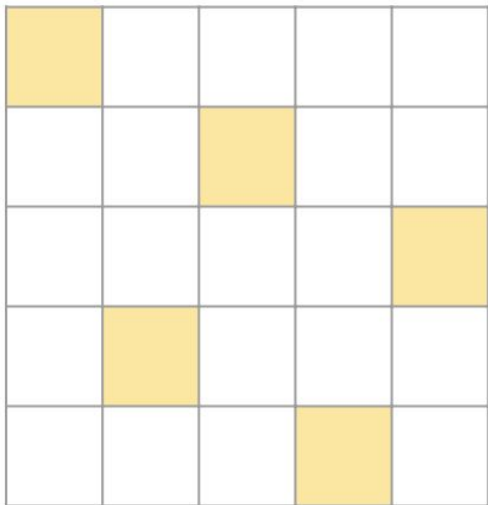
Pruning



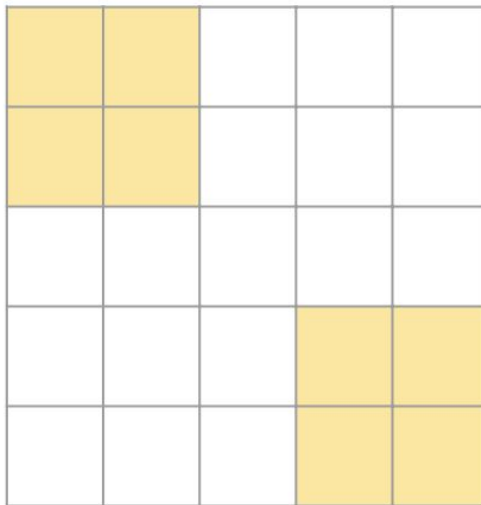




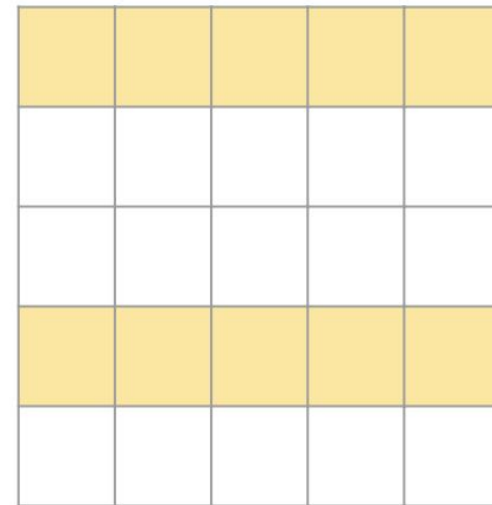
Виды прореживания



Unstructured

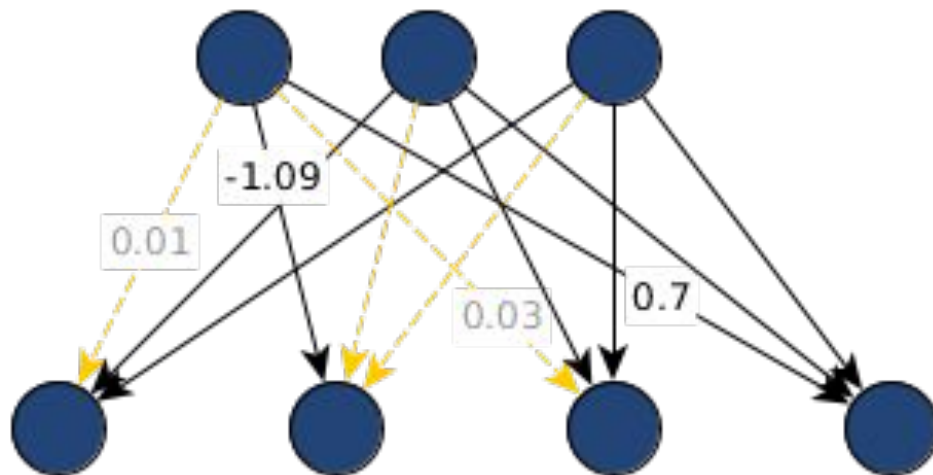


Blocked

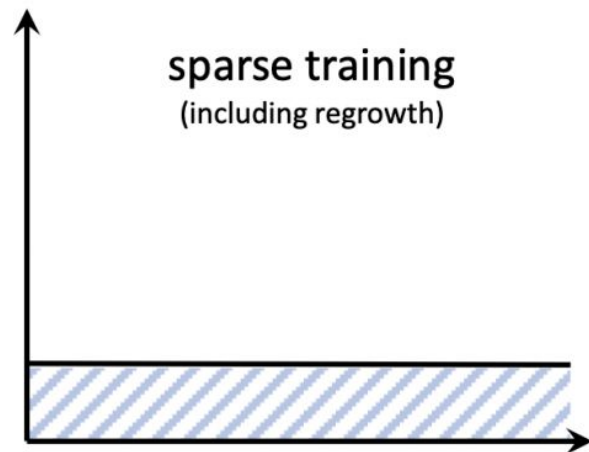
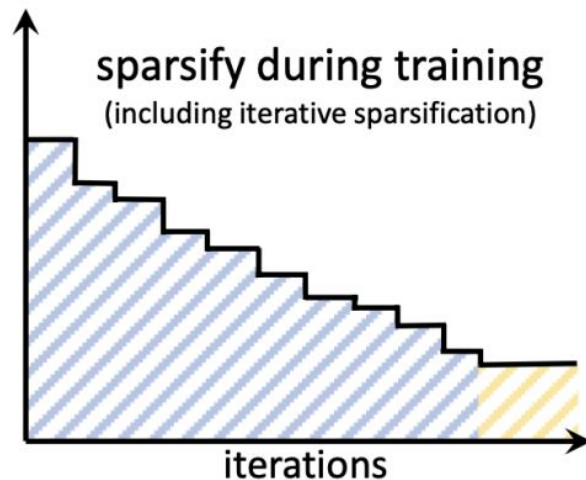
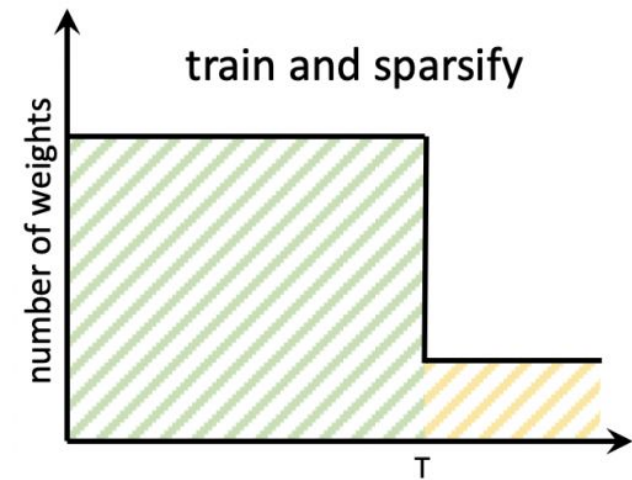


Axis-wise

Magnitude-based pruning

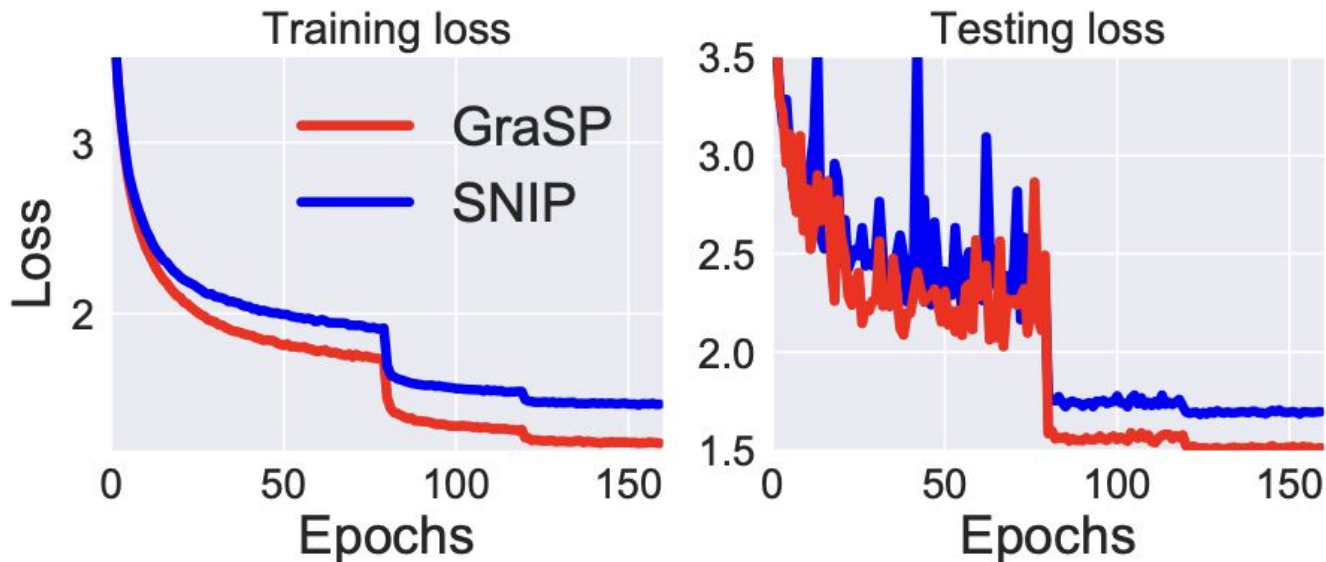


Стратегии прореживания



Sparse training

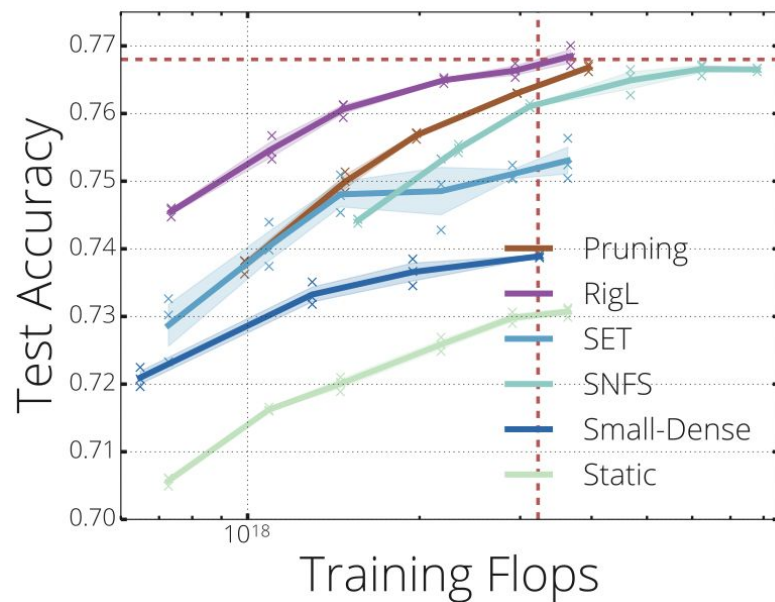
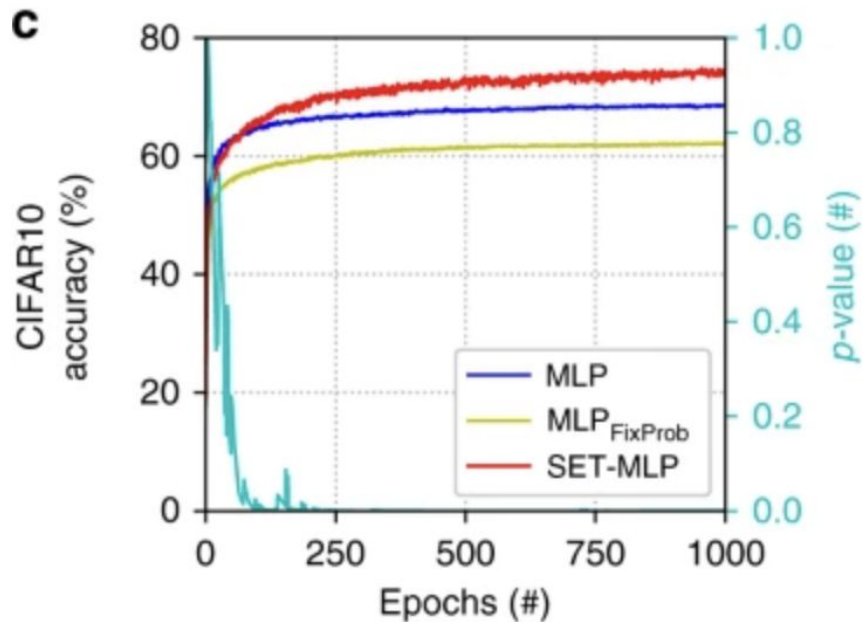
Фиксированная разреженность - фиксируем маску до начала работы GraSP (Gradient-based Sampling with Probabilities), SNIP (Structured Norm Importance Pruning)



Sparse training

Динамическая разреженность - маски постепенно обновляются
(Rigging the Lottery) и SET (Sparse Evolutionary Training)

RigL



Сформулируем гипотезу

Если нейронная сеть инициализируется случайным образом, то в ней существует подсеть, которая может дать лучшую оценку за меньшее количество итераций

Актуальность!

Rigging the Lottery: Making All Tickets Winners

THE LOTTERY TICKET HYPOTHESIS:
FINDING SPARSE, TRAINABLE NEURAL NETWORKS

Stabilizing the Lottery Ticket Hypothesis*

DRAWING EARLY-BIRD TICKETS: TOWARDS MORE EF-
FICIENT TRAINING OF DEEP NETWORKS

**One ticket to win them all: generalizing lottery ticket
initializations across datasets and optimizers**

Linear Mode Connectivity and the Lottery Ticket Hypothesis

PICKING WINNING TICKETS BEFORE TRAINING
BY PRESERVING GRADIENT FLOW

Пути нахождения лотерейных билетов

Iterative pruning

Random reinitialization

One-shot pruning

Различные алгоритмы

1. Берем нейронную сеть $f(x; \theta_0)$
2. Обучаем на протяжении j итераций, у нас сеть с новыми параметрами
3. Прореживаем $p\%$ параметров, создаем маску ненулевых весов
4. Наш “билет” выглядит вот так $f(x; m \odot \theta_0)$

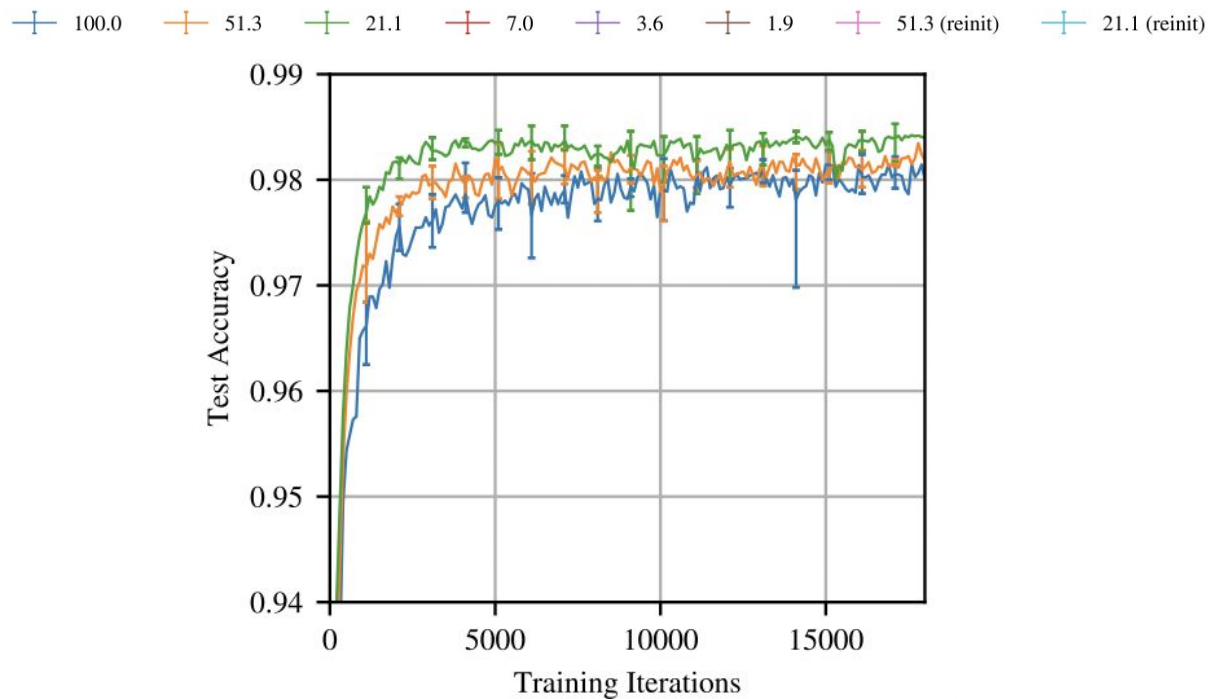
Это one-shot pruning -

Итеративный подход

1. У нас n раундов
2. Каждый раунд удаляется $p^{\frac{1}{n}}\%$ весов, которые остались после прошлого раунда
3. Такой подход эффективнее

Пример - Полносвязные нейронные сети

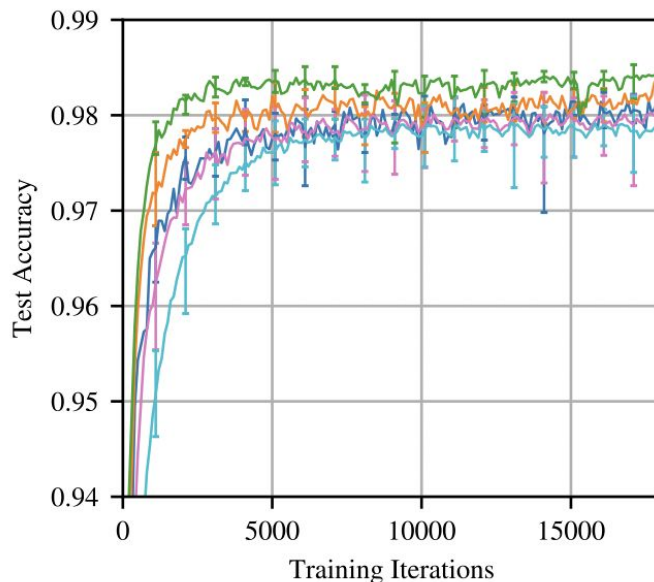
Lenet, iterative pruning



Рандомная реинициализация

В момент, когда мы применяем маску, мы берем не изначальную инициализацию, а рандомную. Работает дольше, менее качественно

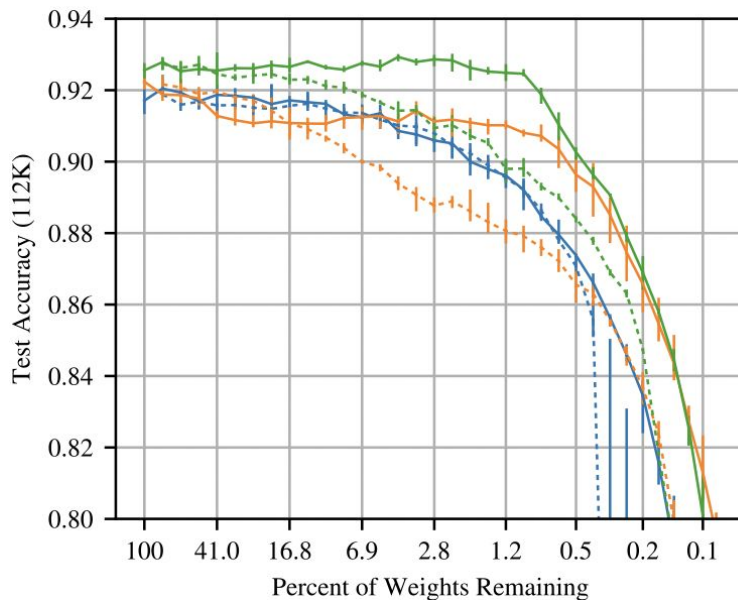
— 100.0 — 51.3 — 21.1 — 7.0 — 3.6 — 1.9 — 51.3 (reinit) — 21.1 (reinit)



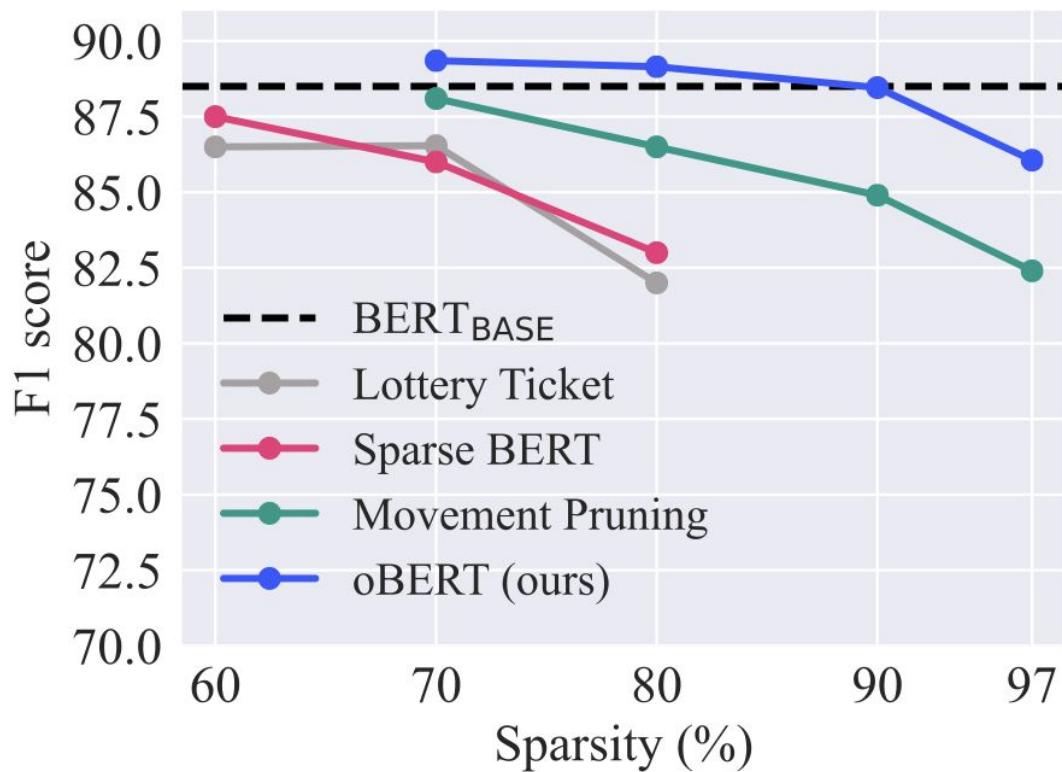
Работа с серьезными датасетами

Чтобы добиться хоть какого-то результата на VGG-19, пришлось поиграться с learning rate, сделать global pruning, добавить warmup

—+— rate 0.1 -.-+.-.- rand reinit —+— rate 0.01 -.-+.-.- rand reinit —+— rate 0.1, warmup 10K -.-+.-.- rand reinit



The optimal BERT Surgeon



Итоги

- Избыточность нейронных сетей
- Большое поле для исследований, экспериментов
- Все зависит от задачи, скорее всего найдется лучшая альтернатива

Полезные ссылки

- [THE LOTTERY TICKET HYPOTHESIS:
FINDING SPARSE, TRAINABLE NEURAL NETWORKS](#)
- [Stabilizing the Lottery Ticket Hypothesis](#)
- [Linear Mode Connectivity and the Lottery Ticket Hypothesis](#)
- [PICKING WINNING TICKETS BEFORE TRAINING BY PRESERVING GRADIENT FLOW](#)
- [Rigging the Lottery: Making All Tickets Winners](#)
- [SET](#)
- [The Optimal BERT Surgeon: Scalable and Accurate Second-Order Pruning for Large Language Models](#)