



Факультет компьютерных наук
Образовательная программа
“Прикладная математика и информатика”



The Lottery ticket Hypothesis

Выполнил:
Юдин Николай Евгеньевич, БПМИ202

Москва, 2022

1. Что такое прунинг и в чем его недостатки?
2. Lottery Ticket Hypothesis и базовый алгоритм поиска билетов
3. Эксперименты и возможные вариации алгоритма
4. Важность начальных данных
5. Применение в других задачах
6. Выводы

Что такое прунинг?

Большие и глубокие
нейросети



Много параметров,
много памяти



Долго обучаемся

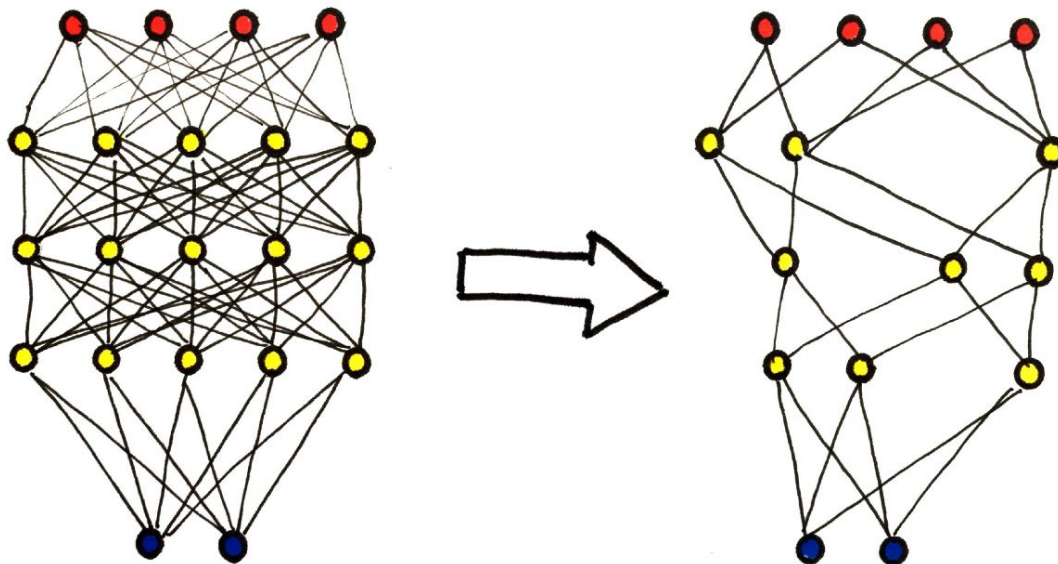


Много времени



Что такое прунинг?

4



Хотим: не обучать большую нейросеть, а потом выкидывать параметры, хочется как-то сразу, за одно обучение нейросети

Основная гипотеза

Пусть дана нейросеть $f(x, \theta)$ с инициализацией θ_0 .

Тогда существует подсеть ($m \in \{0,1\}^{|\theta|}$) такая, что если учить с нуля нейросеть $f(x, \theta \odot m)$ с инициализацией $\theta_0 \odot m$, то

1. Минимум на валидации будет достигнут за меньшее число итераций нежели в исходном случае.
2. Ошибка на валидации будет не больше чем у исходной нейросети
3. Количество параметров у такой подсети сильно меньше чем у исходной

Такую подсеть и будем называть лотерейным билетом.



А как искать билеты?

Базовый алгоритм поиска билетов:

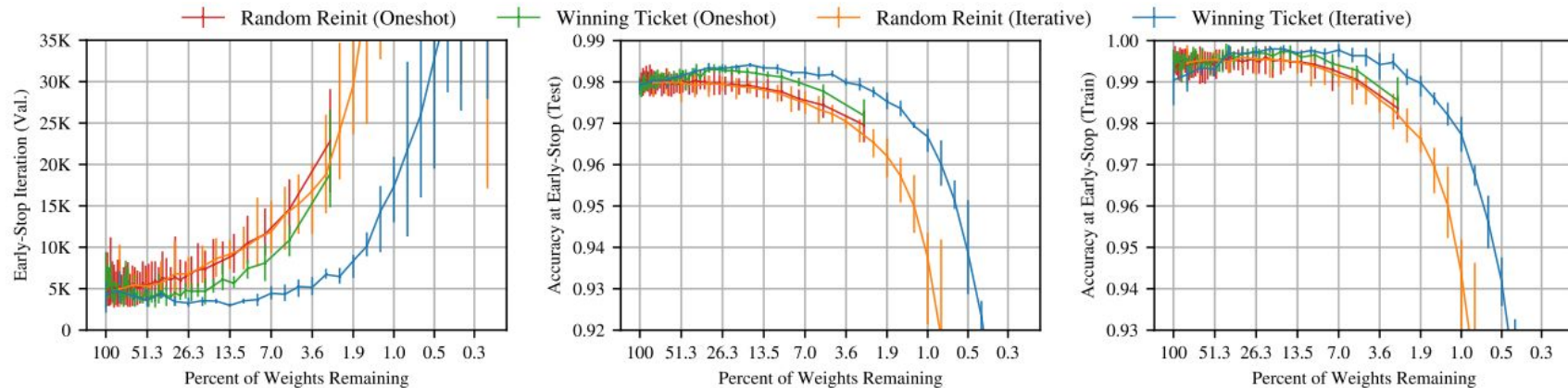
1. Произвольно проинициализируем нейросеть $f(x; \theta_0)$
2. Поучимся какое-то количество итераций, получим новые параметры
3. Сделаем прунинг, проредим какую-то долю, $p\%$ параметров
4. Теперь будем обучать нейросеть $f(x, \theta_0 \odot m)$. Данная нейросеть и есть наш выигрышный билет.

Рубрика “Эксперименты”

7

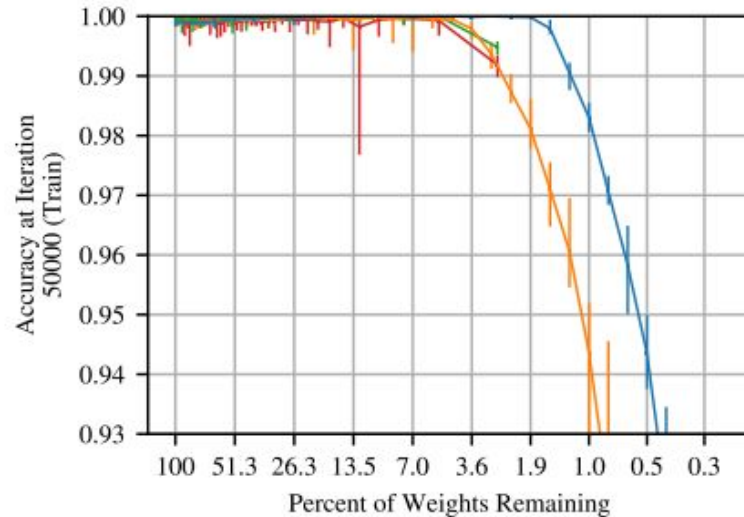
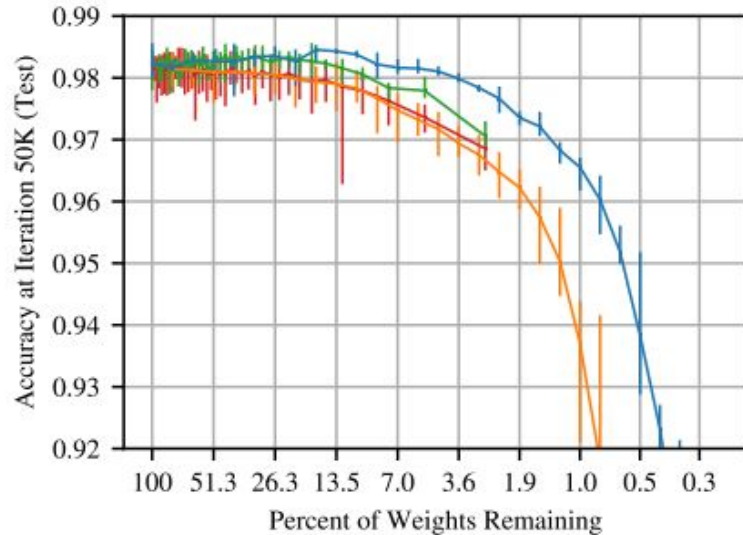
Network	Lenet	Conv-2	Conv-4	Conv-6	Resnet-18	VGG-19
				64, 64, pool	16, 3x[16, 16]	2x64 pool 2x128
			64, 64, pool	128, 128, pool	3x[32, 32]	pool, 4x256, pool
Convolutions		64, 64, pool	128, 128, pool	256, 256, pool	3x[64, 64]	4x512, pool, 4x512
FC Layers	300, 100, 10	256, 256, 10	256, 256, 10	256, 256, 10	avg-pool, 10	avg-pool, 10
All/Conv Weights	266K	4.3M / 38K	2.4M / 260K	1.7M / 1.1M	274K / 270K	20.0M
Iterations/Batch	50K / 60	20K / 60	25K / 60	30K / 60	30K / 128	112K / 64
Optimizer	Adam 1.2e-3	Adam 2e-4	Adam 3e-4	Adam 3e-4	← SGD 0.1-0.01-0.001 Momentum 0.9 →	
Pruning Rate	fc20%	conv10% fc20%	conv10% fc20%	conv15% fc20%	conv20% fc0%	conv20% fc0%

Датасеты: MNIST и CIFAR10



Видно, что итеративный прунинг очень хорош в сравнении с остальными методами.

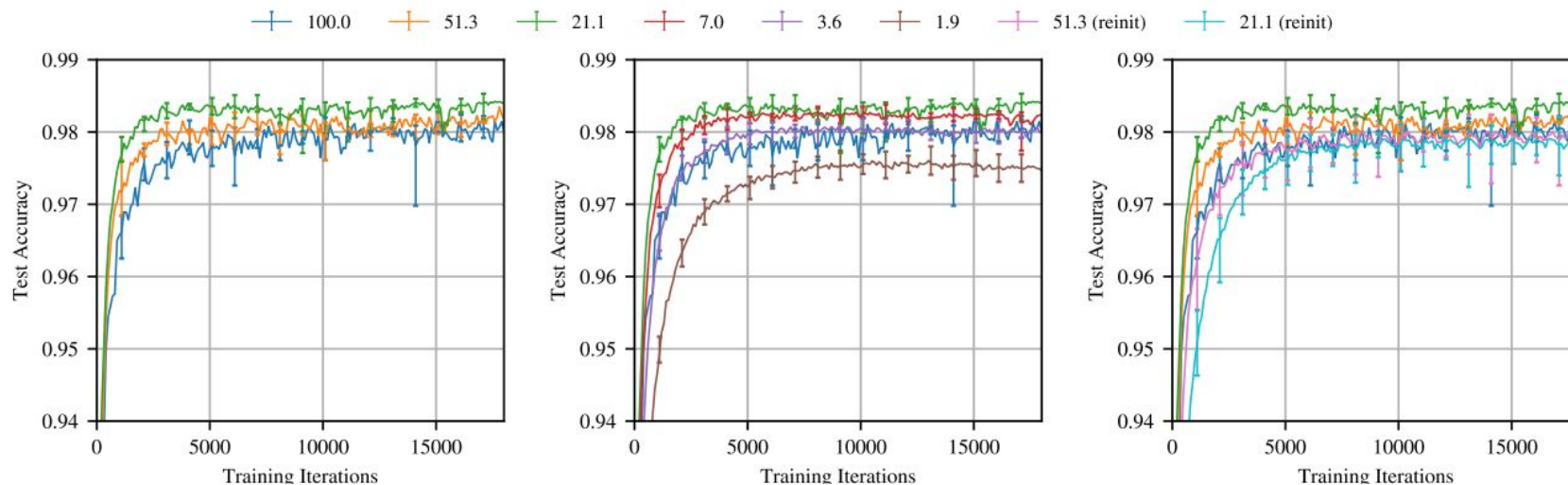
Интересное замечание



Здорово, что лотерейные билеты имеют хорошую обобщающую способность

Рандомно инициализировать плохо?

10

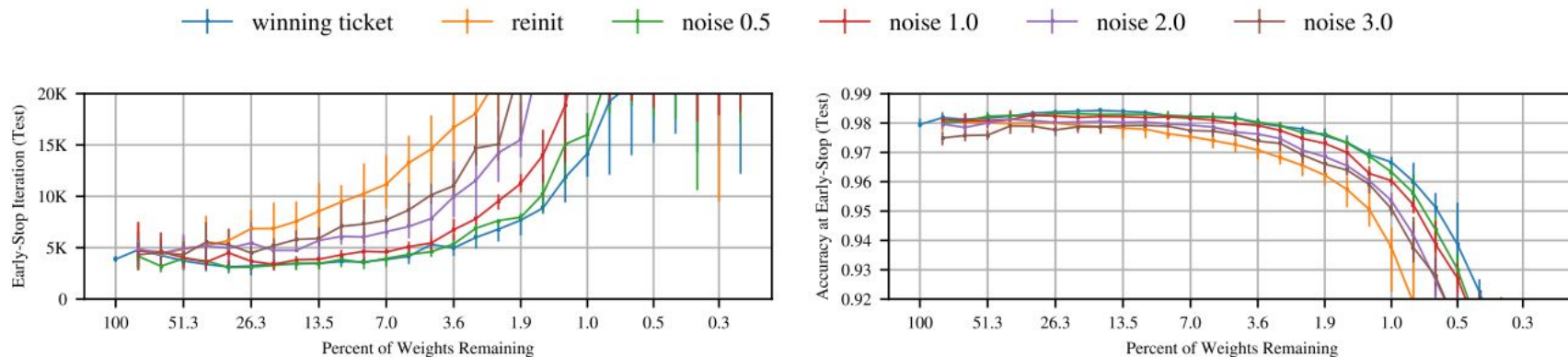


Не стоит рандомно переинициализироваться, обучается долго и плохо

А не совсем рандомно?

11

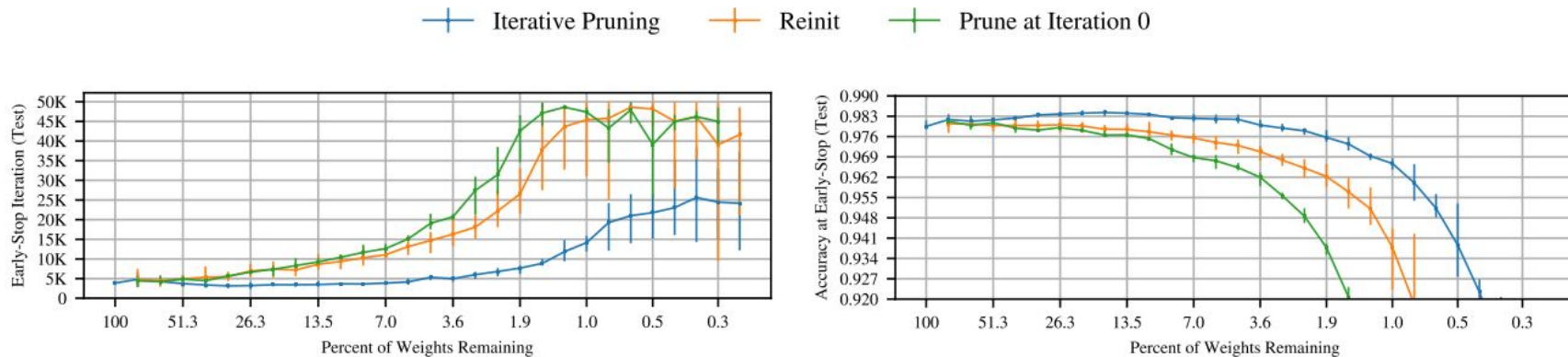
Добавим случайный шум в данные, посмотрим, что получится:



Случайный шум в целом не принес успеха, но и не сильно ухудшил

А если сразу избавиться от лишнего?

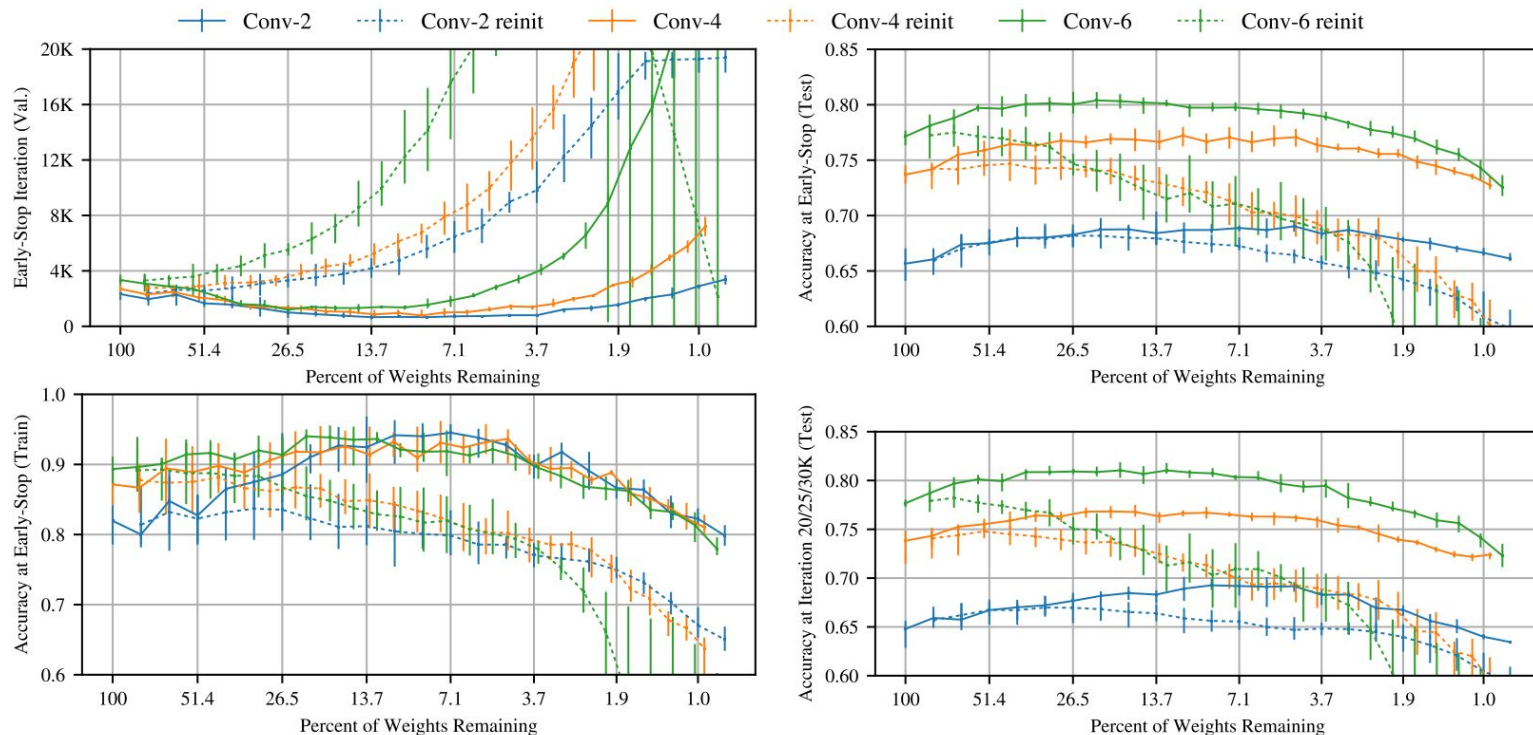
12



Совсем все плохо если убрать без разбора некоторые параметры

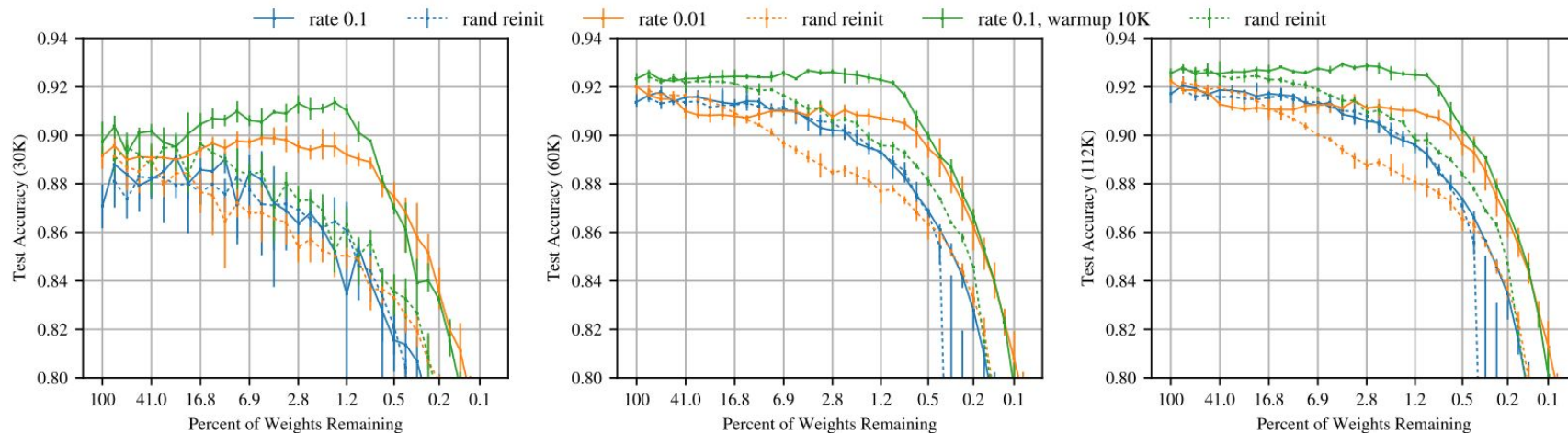
Немного про сверточные нейросети

13



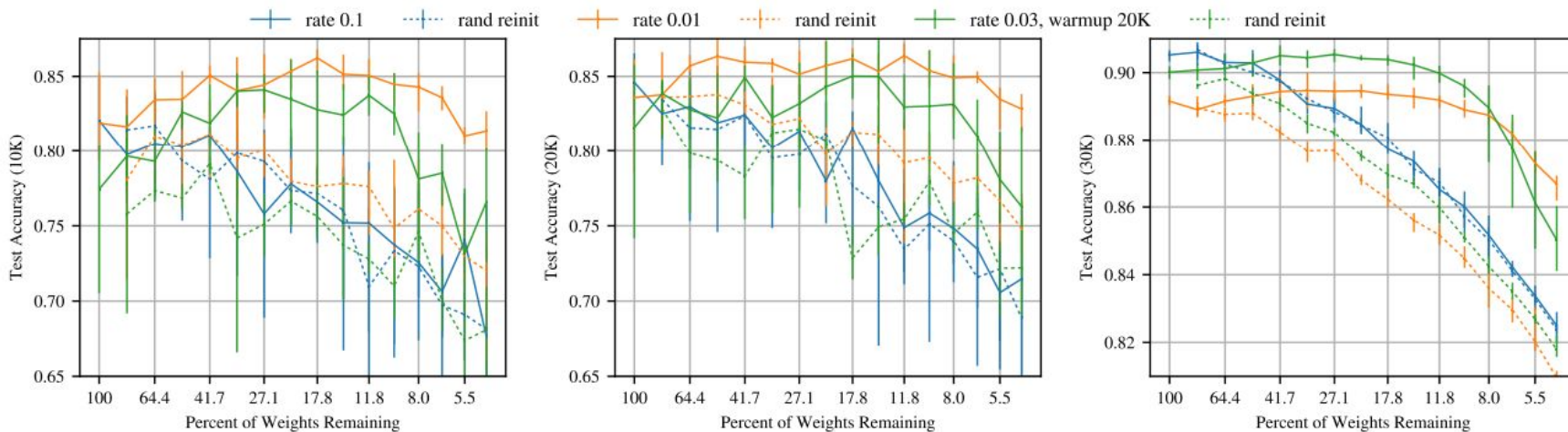
Общая тенденция сохранилась, обучаемся быстрее и не сильно хуже

Обучение VGG19



Проблемы с поиском: при высоком lr появились проблемы, да и разогрев не помог особо

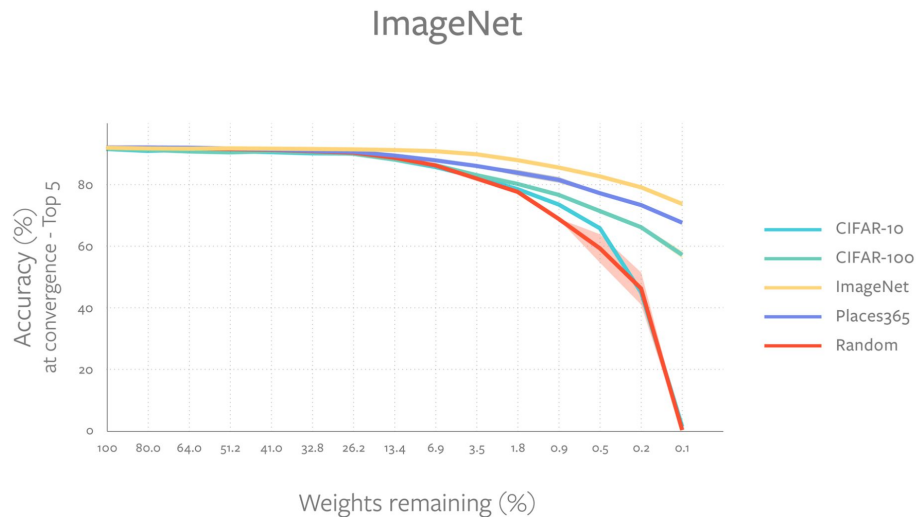
ResNet-18 и его обучение



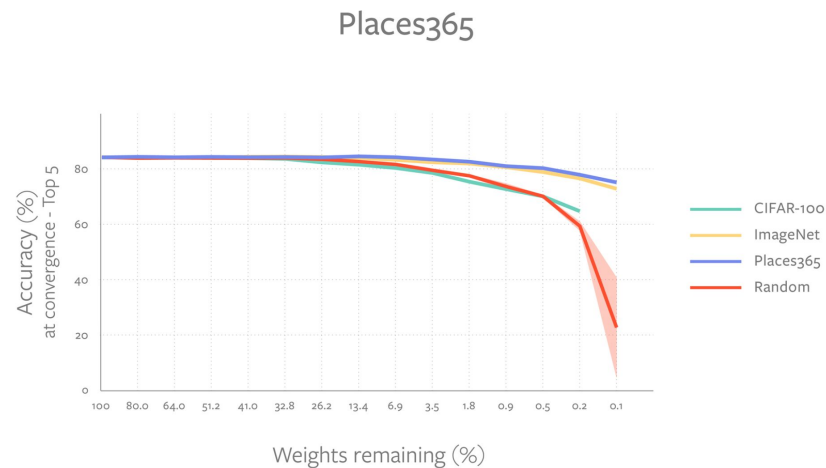
Здесь проблемы с поиском проявляются еще сильнее чем раньше

А что если разные датасеты?!

16



Проверка лотерейного билета
обученного на ImageNet на
остальных датасетах

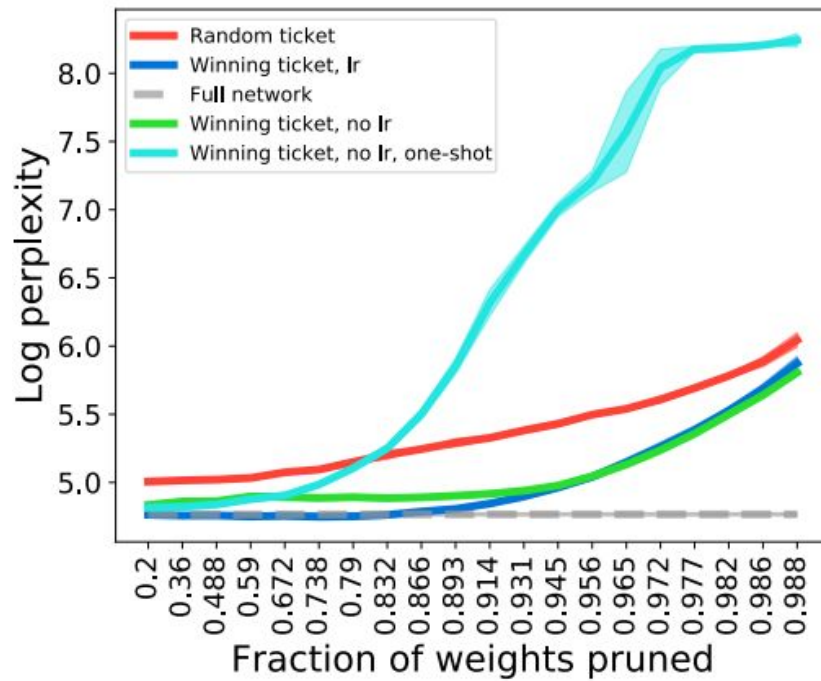


Проверка лотерейного билета
обученного на Places365 на остальных
датасетах.

А что если другая задача?

17

Производительность лотерейных билетов для LSTM обученной на WikiText-2

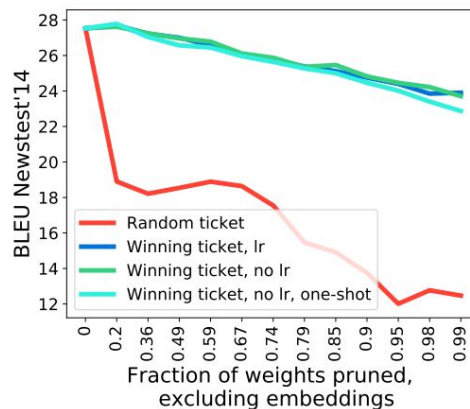
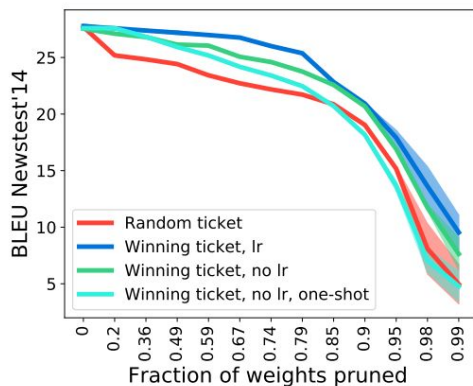


А что если другая задача?

18



Производительность лотерейных билетов для Transformers, обученной на WikiText-2



Поздняя перемотка лучше чем алгоритм из задач классификации

Лотерейные билеты хороши потому что:

- Обучаются действительно быстрее без потери качества
- Переносятся между разными датасетами
- Могут не только в классификацию (надеемся)

Однако

- На высоких l_1 иногда все равно провал
- Иногда тяжело их оптимизировать (некоторые люди и статьи утверждают обратное)

Спасибо за внимание!

<https://arxiv.org/pdf/1803.03635.pdf>

<https://ai.facebook.com/blog/understanding-the-generalization-of-lottery-tickets-in-neural-networks/>

<https://arxiv.org/pdf/1906.02768.pdf>