

1. Loss Surfaces, Mode Connectivity, and Fast Ensembling of DNNs

- 1.1. Какое свойство пространства параметров нейронной сети авторам удалось пронаблюдать в статье?
- 1.2. Алгоритм поиска кривой с низким значением функции потерь: какой функционал оптимизируется, от чего он зависит, что представляет собой итерация поиска кривой?
- 1.3. Поиск среди кривых простого вида (ломаная из двух сегментов): как это позволяет упростить функцию потерь?
- 1.4. Описание алгоритма Fast Geometric Ensembling.

2. The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

- 2.1. Сформулируйте гипотезу лотерейного билета ("the lottery ticket hypothesis"). Что такое выигрышный билет ("winning ticket") в контексте гипотезы?
- 2.2. Опишите предложенный в статье алгоритм поиска выигрышного билета.
- 2.3. Как ведут себя скорость обучения и точность при случайной инициализации весов выигрышного билета и при переиспользовании стартовых весов?

3. Understanding Deep Learning Requires Rethinking Generalization

- 3.1. Что такое ошибка обобщения (generalization error) модели? Какая будет ошибка обобщения у модели AlexNet, если метки изображений MNIST заменить на случайные?
- 3.2. Приведите примеры традиционных методов описания generalization error
- 3.3. Что такое явная и неявная регуляризации? Приведите примеры обоих способов регуляризации
- 3.4. Сколько параметров достаточно двухслойной полносвязной сети с ReLU-активациями для описания функции ответов для n сэмплов размерности d ?

4. Deep Equilibrium Models

- 4.1. Сформулируйте теорему о неявной функции для двумерного случая.
- 4.2. Выведите формулу для градиента функции потерь по весам слоя.
- 4.3. Как представить многослойную нейронную сеть с ReLU активациями в виде одного неявного слоя?
- 4.4. Что делается во время прямого и обратного распространения ошибки в DEQ? В чем отличие от обычных моделей?

5. Neural Ordinary Differential Equations

- 5.1. В статье авторы проводят аналогию между последовательностью вычислений сети ResNet и применением схемы Эйлера для решения дифференциальных уравнений к некоторому дифференциальному уравнению. Что это за уравнение? Ответ поясните
- 5.2. В чем заключаются преимущества Neural ODE по сравнению с ResNet?
- 5.3. Какие ограничения традиционно встречаются у нормализующих потоков? Как их позволяет обойти использование непрерывных нормализующих потоков, основанных на Neural ODE?