

ИССЛЕДОВАНИЕ ГЕОМЕТРИЧЕСКИХ СВОЙСТВ ФУНКЦИЙ ПОТЕРЬ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ

Гарипов Тимур Исмагилевич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: timgaripov@gmail.com

Одними из наиболее известных параметрических методов машинного обучения являются методы, основанные на нейронных сетях. Настройка параметров нейронной сети осуществляется путем минимизации некоторой функции потерь, которая определяется решаемой задачей. Функции потерь глубоких нейронных сетей имеют сложную структуру, они невыпуклы и могут зависеть от миллионов параметров. Геометрические свойства поверхностей уровня таких функций до сих пор недостаточно исследованы. Известно, что даже для простых нейронных сетей число локальных оптимумов и седловых точек функции потерь очень велико и экспоненциально увеличивается с размерностью пространства параметров [1, 2]. Кроме того, функция потерь принимает большие значения на прямолинейном отрезке, соединяющем два оптимума [3, 4]. Эти наблюдения приводят к предположению об изолированности локальных оптимумов.

В данной работе предложен алгоритм, находящий непрерывную кривую в пространстве параметров (весов) нейронной сети, соединяющую два заданных локальных оптимума, такую, что для всех точек кривой значение функции потерь незначительно отличается от оптимального. Экспериментально показано, что для широкого класса архитектур нейронных сетей такие кривые можно найти в виде кривой безье второго порядка или ломанной линии из двух звеньев.

На изображении (Рис. 1.) в трёх различных плоскостях проиллюстрированы линии уровня кросс-энтропийной функции потерь с L_2 регуляризацией для сети ResNet-164 на задаче классификации изображений из набора данных CIFAR-100. На левом изображении функция потерь изображена в плоскости, содержащей все аффинные комбинации весов трёх нейронных сетей, независимо обученных из случайных начальных приближений. В этой плоскости все оптимумы изолированы. Центральное и правое изображения демонстрируют функцию потерь в плоскостях, содержащих две кривые, найденные предложенным алгоритмом. Крайними точками кривых являются две независимо обученные нейронные сети, соответствующие

оптимумам в нижней части изображения слева.

В рамках данной работы:

- Предложен алгоритм поиска непрерывной кривой между локальными оптимумами, в точках которой значения функции потерь малы.
- Проведены эксперименты, демонстрирующие возможность нахождения описанных кривых для разнообразных архитектур нейронных сетей и наборов данных.
- На основе полученных результатов был предложен новый метод построения ансамблей неронных сетей, позволяющий получить ансамбль за время, необходимое для обучения одной сети. Результаты экспериментального исследования показали, что предложенный метод превосходит конкурирующий подход [5].

Полученные в данной работе результаты могут оказать влияние на исследования в области нейронных сетей и найти свое применение в таких направлениях как: повышение эффективности и точности обучения, создание ансамблей нейронных сетей, выбор семейства распределений для аппроксимации апостериорного распределения в байесовских методах глубинного обучения.

Иллюстрации

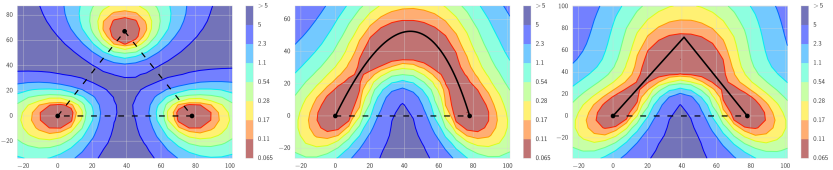


Рис. 1. Визуализация линий уровня функции потерь нейронной сети в плоскостях, содержащих: веса трех независимо обученных сетей (слева); кривую безье (по центру) и ломанную линию (справа), соединяющие две точки, расположенные в нижней части на графике слева.

Литература

1. Dauphin, Yann N, Pascanu, Razvan, Gulcehre, Caglar, Cho, Kyunghyun, Ganguli, Surya, and Bengio, Yoshua. Identifying and attacking the saddle point problem in high-dimensional non-convex

- optimization. // In Advances in neural information processing systems, pp. 2933–2941, 2014.
2. Choromanska, Anna, Henaff, Mikael, Mathieu, Michael, Arous, Gerard Ben, and LeCun, Yann. The loss surfaces of multilayer networks. // In Artificial Intelligence and Statistics, pp. 192–204, 2015.
 3. Goodfellow, Ian J, Vinyals, Oriol, and Saxe, Andrew M. Qualitatively characterizing neural network optimization problems. // International Conference on Learning Representations, 2015.
 4. Keskar, Nitish Shirish, Mudigere, Dheevatsa, Nocedal, Jorge, Smelyanskiy, Mikhail, and Tang, Ping Tak Peter. On large-batch training for deep learning: Generalization gap and sharp minima. // International Conference on Learning Representations, 2017.
 5. Huang, Gao, Li, Yixuan, Pleiss, Geoff, Liu, Zhuang, Hopcroft, John E, and Weinberger, Kilian Q. Snapshot ensembles: Train 1, get m for free. // arXiv preprint arXiv:1704.00109, 2017.