

Автоматическое построение нейросети оптимальной сложности

Маркин В. О.

markin1198@mail.ru

Московский физико-технический институт

В работе рассматривается задача построения оптимальной структуры нейронной сети и исследуется вопрос устойчивости построенной модели. Для оптимизации структурных параметров используется переход от выбора конкретной архитектуры к выбору комбинации различных архитектур сети и вариационный подход. Также исследуется влияние изменения данных на структуру сети. Для оценки качества и устойчивости моделей, построенных при помощи данного метода, проводятся эксперименты на выборке Boston, MNIST и синтетических данных. Проводится сравнение предложенного алгоритма с другими методами поиска оптимальных моделей нейронной сети.

Ключевые слова: *нейронные сети, оптимизация гиперпараметров, устойчивость нейросетевой модели.*

1 Введение

При использовании нейростетевых моделей в анализе данных часто встает вопрос о выборе архитектуры модели. Нейронная сеть имеет большое число гиперпараметров и долгое время - для выбора архитектуры использовался перебор и различные эвристические соображения. Такой подход вычислительно неэффективен и не дает гарантий оптимальности полученной модели.

В данной работе рассматривается задача построения оптимальной нейронной сети. Ожидается, что оптимальная модель будет не только давать хорошее качество, но и иметь максимально простую структуру. Под структурой нейронной сети понимается набор гиперпараметров, таких как число слоев, размерность каждого слоя, функции активации и параметры регуляризации.

Для выбора оптимального набора гиперпараметров проводится процедура релаксации(*) - переход от дискретного множества возможных значений гиперпараметров к непрерывному множеству их комбинаций. Эта процедура позволяет параметризовать структуру модели некоторым действительным вектором. Такой подход дает возможность применять различные методы оптимизации для нахождения наилучшего набора гиперпараметров. Оптимизация гиперпараметров проводится градиентными методами (*) либо с использованием Гауссовских процессов и Байесовской оптимизации.

Проведены вычислительные эксперименты на выборке Boston, MNIST и синтетических данных. В экспериментах оценивается не только качество, которое дает полученная модель, но и её вычислительная сложность и устойчивость. Проведено сравнение различных алгоритмов построения оптимальной модели.