

# СОВЕРШЕНСТВОВАНИЕ ПРОЦЕДУР ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Филанович К. Н.

Железко Б. А. - канд. техн. наук, доцент

Процесс обучения нейронных сетей является одним из самых сложных в предметной области. Для его упрощения необходимо иметь чёткое представление о том, как те или иные параметры модели нейронной сети будут влиять на её эффективность. В данной работе будет произведен анализ ключевых параметров модели нейронной сети, позволяющей совершенствовать процедуры её обучения.

Совершенствование процедур обучения нейронных сетей ставит своей целью ускорение и улучшение адапционных механизмов установления весов модели в ответ на набор тренировочных данных. Процесс улучшения механизмов адаптации должен начинаться с тщательной конфигурации основных параметров модели нейронной сети с задействованием, например, стохастического градиентного спуска и выставления весов с использованием алгоритма обратного распространения ошибки. Ниже каждый из вопросов будет рассмотрен более подробно.

Выбор наиболее подходящего градиентного спуска. Эффективность модели напрямую зависит от того, какая функция градиентного спуска была выбрана: как пакетная, стохастическая или мини-пакетный градиентный спуск;

1) Настройка функции потерь. Понимание того, как различные функции потерь интерпретируют данные, может ли альтернативная функция эффективно применима к рассматриваемой предметной области [1].

Стохастический градиентный спуск — это общий алгоритм оптимизации, который можно применять к широкому кругу задач предметной области. Тем не менее, процесс оптимизации (или процесс обучения) может стать нестабильным, и могут потребоваться дополнительные действия:

1) «Исчезающие» градиенты. Не допускают обучение глубоких уровней, в результате чего уровни, близкие к входному слою, не будут обновлять свои веса. Решение этой проблемы происходит с помощью современных функций активации, таких как, например, выпрямленная функция линейной активации.

«Взрывающиеся» градиенты. Значительное обновление весов приводят к числовому переполнению или наоборот, исчезновению порядка, в результате чего веса сети принимают значение «NaN» или «Inf», подобное может быть предотвращено с использованием градиентного масштабирования или градиентного отсечения [2].

Ограничение набора данных по некоторым задачам прогнозного моделирования может помешать эффективному обучению. Для ускорения процесса оптимизации можно использовать специализированные методы, предоставляя полезный начальный набор весов или даже целые модели, которые можно использовать для извлечения признаков; например:

1) «Жадная» многоуровневая предварительная подготовка. В местах добавления в модель новых слоёв учат интерпретировать результаты предыдущих уровней что позволяет разрабатывать гораздо более глубокие модели нейронных сетей.

Обучение на примере. В тех случаях когда уже существует обученная модель по другой, но каким-то образом связанной с рассматриваемой задаче прогнозного моделирования можно использовать её для определения весов или использовать её в качестве модели извлечения входных значений для предоставления входных данных для будущей модели сети [3].

На основании всего вышесказанного можно с уверенностью сказать, что процесс совершенствования процедур обучения нейронных сетей включает в себя множество подпроцессов, каждый из которых так или иначе оказывает существенное влияние на итоговую гибкость модели нейронной сети и её эффективность. Лишь комплексный подход к вопросу её обучения может гарантировать успешность применимости нейронной сети в контексте предметной области.

## Список использованных источников:

1. Reed, R. Neural Smoothing: Supervised Learning in Feedforward Artificial Neural Networks / R. Reed, Robert J Marks. – Massachusetts : A Bradford Book, 1999. – 358 S.
2. Bishop M., Christopher. Pattern Recognition and Machine Learning Christopher M. Bishop. – Cambridge : Springer Science+Business Media, 2011. – 738 S.
3. Goodfellow I., Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – Massachusetts : The MIT Press, 2016. – 775 S.