

УДК 004.855.5

ПРОБЛЕМАТИКА МАШИННОГО ОБУЧЕНИЯ: ВЛИЯНИЕ ГИПЕРПАРАМЕТРОВ, ПЕРЕОБУЧЕНИЕ И НЕДООБУЧЕНИЕ НЕЙРОННЫХ СЕТЕЙ



А.А. Навроцкий

Заведующий кафедрой
информационных технологий
автоматизированных систем
БГУИР, кандидат физико-
математических наук, доцент
navrotsky@bsuir.by



А.О. Шведко

Аспирант кафедры
информационных
технологий
автоматизированных
систем БГУИР, магистр
kv.space@yandex.ru



В.И. Сакович

Научный сотрудник, магистр
kv.space@yandex.ru

А.А. Навроцкий

Заведующий кафедрой информационных технологий автоматизированных систем БГУИР,
кандидат физико-математических наук, доцент.

А.О. Шведко

Окончил Белорусский государственный университет информатики и радиоэлектроники. Магистр.

В.И. Сакович

Научный сотрудник, магистр.

Аннотация. В рамках настоящего исследования выполнен анализ влияния входных данных, параметров первоначальной настройки на процесс переобучения и недообучения нейронной сети. Это исследование имеет важное значение, поскольку первоначальная настройка модели обучения является ключевыми аспектом в процессе обучения моделей глубокого обучения.

В процессе анализа модели были обучены с использованием различных комбинаций количества нейронов с одинаковыми входными данными и алгоритмом оптимизации, что позволило получить информацию о том, как обучение влияет на тренировочную и валидационную потери.

Ключевые слова: TensorFlow, гиперпараметры, переобучение и недообучение, нейронные сети, машинное обучение.

Введение. Наиболее распространенными программными библиотеками для машинного обучения являются *TensorFlow* и *PyTorch*. *TensorFlow*, разработанный компанией *Google*, известен своей мощной экосистемой, включающей инструменты для развертывания моделей и их оптимизации. *PyTorch*, созданный *Facebook*, выделяется интуитивно понятным интерфейсом и гибкостью, что делает его предпочтительным выбором для исследователей. Обе библиотеки активно поддерживаются и развиваются, предоставляя пользователям доступ к последним достижениям в области машинного обучения, а их популярность в академическом и промышленном секторах подтверждает эффективность и надежность.

Однако, процесс обучения нейронных сетей сопряжен с рядом значительных трудностей, которые могут существенно повлиять на качество и надежность получаемых моделей. Эти трудности включают в себя переобучение, недообучение и грамотное использование гиперпараметров, а их преодоление является важной задачей для

исследователей, стремящихся улучшить производительность и обобщающую способность моделей [1].

Влияние гиперпараметров на обучение модели. Гиперпараметры играют ключевую роль в процессе обучения нейронных сетей, поскольку они определяют стратегию оптимизации и влияют на динамику обучения, производительность и способность модели к обобщению.

Конкретные гиперпараметры, такие как скорость обучения, оказывают значительное влияние на производительность модели. Например, эксперименты с *ImageNet* показали, что изменение скорости обучения на порядок может существенно повлиять на скорость сходимости модели. Слишком высокая скорость обучения может привести к нестабильности и невозможности достижения оптимума, в то время как слишком низкая – к медленной сходимости.

Однако, такие параметры, как скорость оптимизации и тип используемого оптимизатора (например, *SGD*, *Adam*, *RMSprop*), также влияют на обучение. Каждый оптимизатор имеет свои особенности и может быть более или менее эффективным для разных задач и архитектур.

Правильная настройка скорости обучения и оптимизации является критически важным этапом в обучении нейронных сетей.

Для наглядности приведенных сведений необходимо построить несколько простых полносвязных сетей (т.е. многослойных перцептронов) со следующими параметрами:

Таблица 1. Параметры конфигурации простых полносвязных сетей

Наименование модели	Модель-основание	Малая модель (недообучение)	Большая модель (переобучение)
Алгоритм оптимизатора	<i>Adam</i>	<i>Adam</i>	<i>Adam</i>
Размер гиперпараметра (<i>batch size</i>)	512	512	512
Количество эпох	20	20	20
Ректифицированный линейный юнит первого слоя (количество нейронов)	16	4	300
Ректифицированный линейный юнит второго слоя (количество нейронов)	16	4	300
Сигмоидная функция (количество нейронов)	1	1	1

Причины переобучения. Переобучение является важной проблемой, с которой сталкиваются при машинном обучении. Оно возникает, когда модель демонстрирует отличные результаты на обучающем наборе данных, но её производительность на тестовых данных значительно ниже. Это указывает на неспособность модели к адаптации, т. е. к применению выявленных закономерностей на новых, ранее не обработанных данных. Таким образом, переобучение свидетельствует о чрезмерной адаптации модели к особенностям обучающего набора, включая информационный шум и случайные вариации.

Основными причинами переобучения являются чрезмерная сложность модели и недостаточное количество обучающих данных. Когда модель имеет слишком большое количество параметров, она может легко запомнить обучающие данные, включая их шум, вместо того, чтобы выучить общие закономерности.

Для наглядного представления создадим график тренировочной и валидационной потери большой модели относительно модели-основания (рисунок 1), где синим цветом представлена модель-основание, а зеленым большая модель. Непрерывные линии показывают тренировочную потерю, а пунктирные линии показывают валидационную потерю (следует помнить, более низкое значение валидационной потери означает лучшее качество модели). Здесь большая сеть начинает переобучаться позже, чем модель-основание и ее производительность уменьшается намного медленней после начала переобучения.

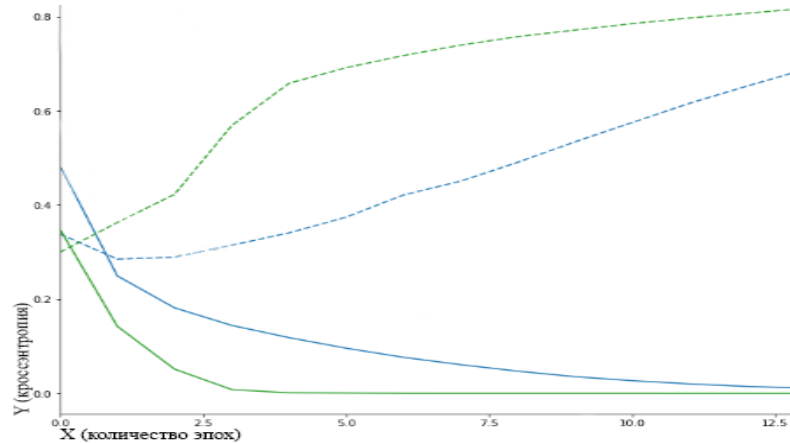


Рисунок 1. Пример тренировочных и валидационных потерь

В этом контексте переобучение можно определить, как ситуацию, «когда ошибка на данных обучения мала, а на данных уточнения значительно больше».

Недообучение и причины возникновения. Недообучение возникает в тех случаях, когда модель нейронной сети не может эффективно обучиться на предоставленных данных, что приводит к низкой производительности как на обучающем, так и на тестовом наборах. Основными причинами этого являются недостаточная сложность архитектуры модели, недостаточное количество эпох обучения и использование неподходящих методов оптимизации, которые не позволяют ей выявить все необходимые закономерности во входных данных.

Как и в случае переобучения модели, создадим график, показывающий недообучение меньшей модели относительно модели-основания (рисунок 2), где синим цветом представлена модель-основание, а оранжевым меньшая модель. Здесь меньшая сеть начинает обучение гораздо раньше, чем модель-основание, однако ее производительность оставляет желать лучшего из-за нехватки входных данных.

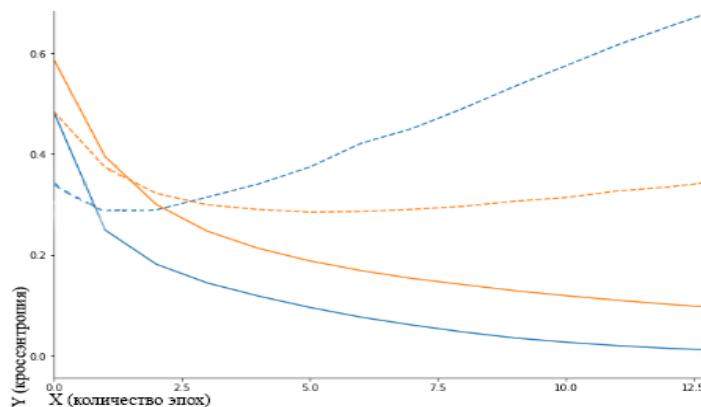


Рисунок 2. Пример тренировочных и валидационных потерь

Недообучение может привести к значительному снижению производительности модели, что выражается в высокой вероятности возникновения погрешности на всех этапах проверки и ограничивает возможность использования модели в реальных приложениях, так как она не способна адекватно анализировать новые входные данные.

Заключение. Широкий спектр задач, решаемых нейронными сетями, делает невозможным создание универсальных и мощных моделей, что, в свою очередь, требует разработки специализированных решений, функционирующих по различным алгоритмам.

Понимание и решение трудностей, связанных с обучением нейронных сетей, является важным этапом в развитии искусственного интеллекта. Перед началом обучения крайне важно провести тщательный анализ данных, который включает проверку их полноты, наличие пропусков и аномалий, а также оценку распределения. Правильная предобработка, включая нормализацию и масштабирование, может значительно улучшить качество модели. Например, пропуски в данных могут быть заполнены средними значениями или предсказаны

с использованием других признаков, что минимизирует их влияние на обучение.

Список литературы

- [1] Шакла Н. Машинное обучение и TensorFlow. СПб.: Питер, 2019. – 336 с.

Авторский вклад

Навроцкий Анатолий Александрович – постановка задачи на исследование и руководство по оценке влияния гиперпараметров на процесс машинного обучения, а также влияние переобучения и недообучения на выходную модель данных.

Шведко Александр Олегович – исследование и оценка доступных материалов по машинному обучению, краткое описание проблематики по теме статьи, построение модели-основания и малой модели.

Сакович Владимир Игоревич – построение большой модели нейронной сети, оформление графического материала, написание заключения.

PROBLEMS OF MACHINE LEARNING: IMPACT OF HYPERPARAMETERS, OVERFITTING AND UNDERFITTING OF NEURAL NETWORKS

A.A. Navrotsky

Head of the Department of
Information Technologies of
Automated Systems BSUIR, PhD in
Physics and Mathematics,
Associate Professor

A.O. Shvedko

Postgraduate student of the
Department of Information
Technologies of Automated
Systems of BSUIR, Master of
Science

V.I. Sakovich

Researcher, Master of Science

Abstract: This study analyzes the influence of input data and initial tuning parameters on the process of overfitting and underfitting of a neural network. This study is important because the initial tuning of the training model is a key aspect in the training process of deep learning models.

In the analysis, the models were trained using different combinations of the number of neurons with the same input data and optimization algorithm, which provided information on how training affects the training and validation losses.

Keywords: TensorFlow, hyperparameters, overfitting and underfitting, neural networks, machine learning.