

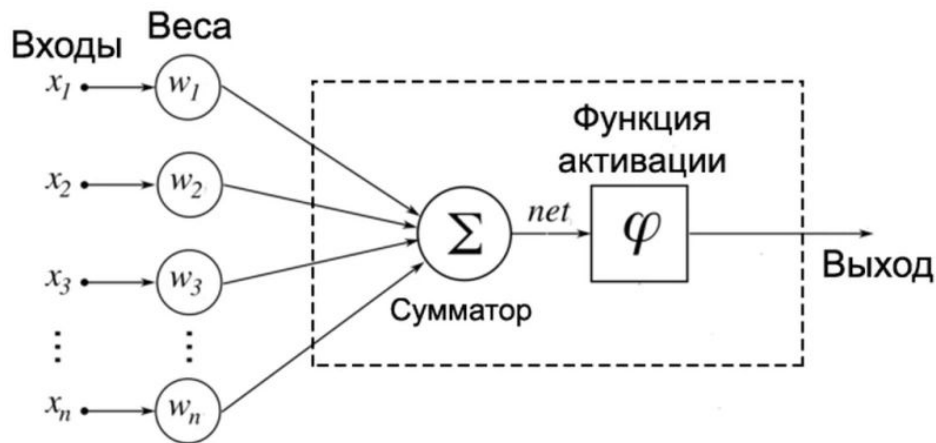
Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт

Оптимизация процесса обучения искусственных нейронных сетей с помощью методов регуляризации весов

Выполнили студенты гр. 5030102/10101
Белоус Ф. В., Ушкарёв С. Д.

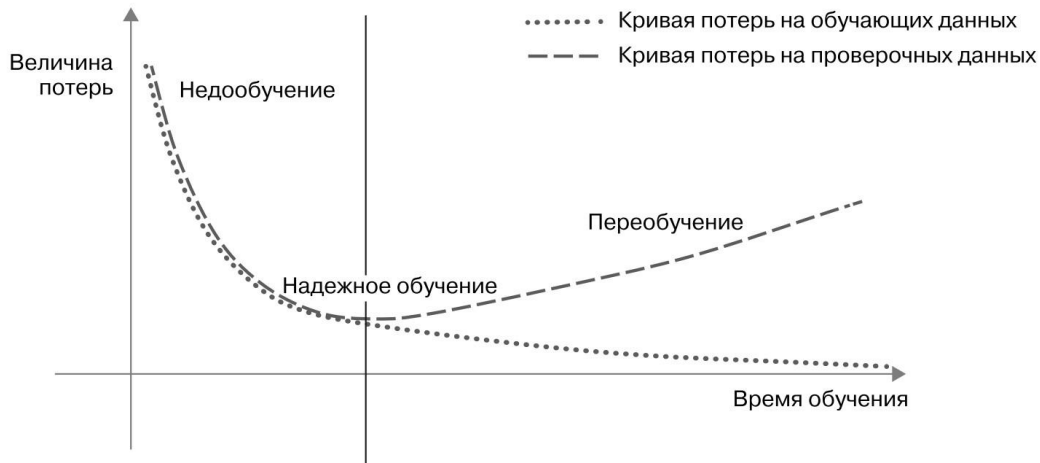
Постановка задачи

- Провести исследование наиболее распространённых методов регуляризации и оценить их эффективность оптимизации процесса обучения нейронных сетей.
- Определить оптимальный набор гиперпараметров для решения задачи из области машинного обучения и сравнить эффективность методов.



Переобучение

Переобучение нейронных сетей – явление, возникающее, когда модель "запоминает" обучающие данные, вместо того, чтобы обобщать закономерности и шаблоны, существующие в данных.



L1 и L2 регуляризация

- **L1 и L2 регуляризация** – методы, основанные на добавлении к функции потерь штрафа на веса модели во время обучения.

$$L1 = \text{Loss} + \lambda \sum_{i=1}^n |w_i|$$

$$L2(y_{pred}, y_{true}, w) = L(y_{pred}, y_{true}) + \lambda \sum_{i=1}^n w_i^2$$

$$\nabla L2(y_{pred}, y_{true}, w) = \nabla L(y_{pred}, y_{true}) + 2\lambda w$$

$$w^{i+1} = w^i - \mu \nabla L2_i(y_{pred}, y_{true}, w_i)$$

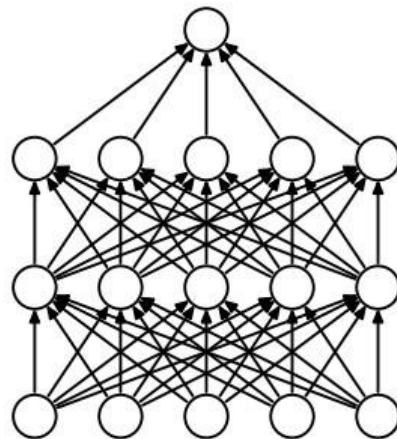
$$w^{i+1} = w^i - \mu (\nabla L(y_{pred}, y_{true}) + 2\lambda w_i)$$

$$w^{i+1} = (1 - 2\mu\lambda)w^i - \mu \nabla L(y_{pred}, y_{true})$$

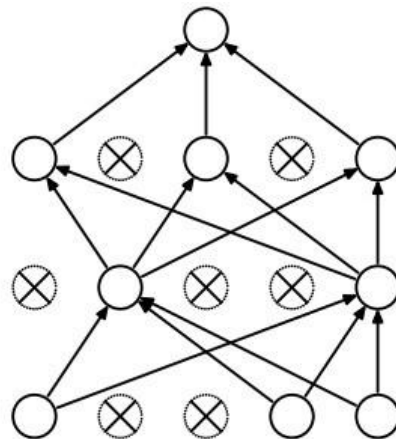
Метод исключения (Dropout)

- **Метод исключения или Dropout** – метод, главной идеей которого является обучение группы из нескольких сетей и усреднение полученных весов вместо обучения одной нейронной сети.

$$O_i = X_i \cdot f\left(\sum_{k=1}^{d_i} w_k x_k + b\right) = \begin{cases} f(\sum_{k=1}^{d_i} w_k x_k + b), & X_i = 1 \\ 0, & X_i = 0 \end{cases}$$



(a) Standard Neural Net



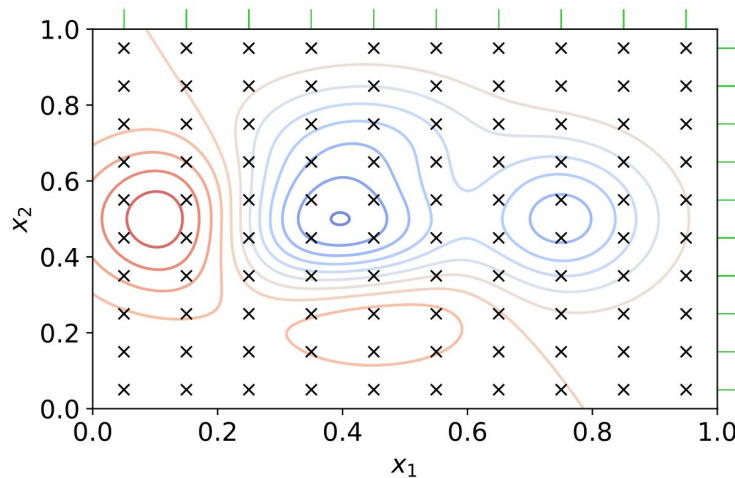
(b) After applying dropout.

Метод оптимизации гиперпараметров

Метод равномерного поиска (Grid Search) – алгоритм поиска оптимальных параметров, заключающийся в итерационном построении равномерной n -мерной сетки в пространстве параметров и вычислении значения целевой функции в узлах сетки для подбора оптимальных значений гиперпараметров.

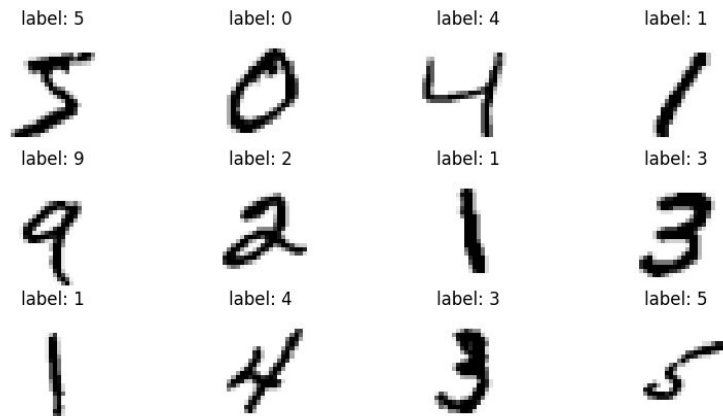
$$\lambda_j = \{\lambda_j^i\} = \left\{a_i + \frac{b_i - a_i}{k}j, i = \overline{1, n}\right\}$$

$$\lambda_{opt}^i \in [\lambda_{j-1}^i, b_i] \text{ и } \lambda_{opt}^i \in [a_i, \lambda_{j+1}^i] \Rightarrow \\ \Rightarrow \lambda_{opt}^i \in [\lambda_{j-1}^i, \lambda_{j+1}^i], i = \overline{1, n}$$



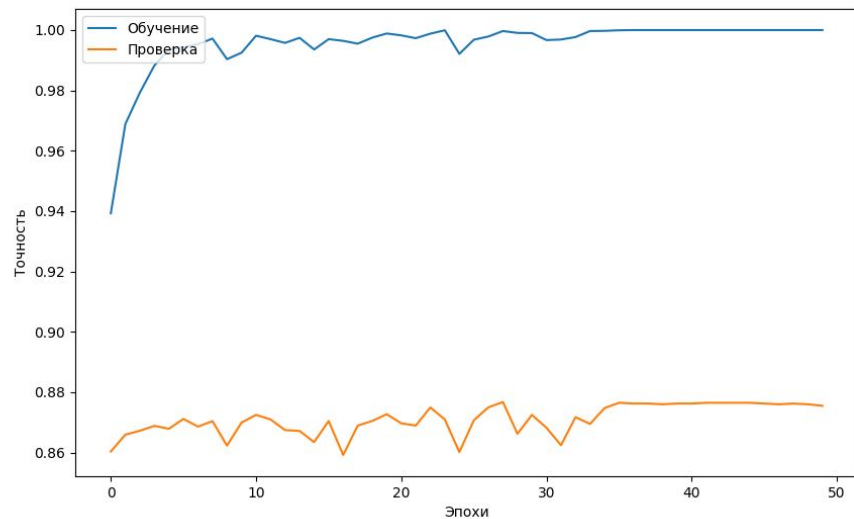
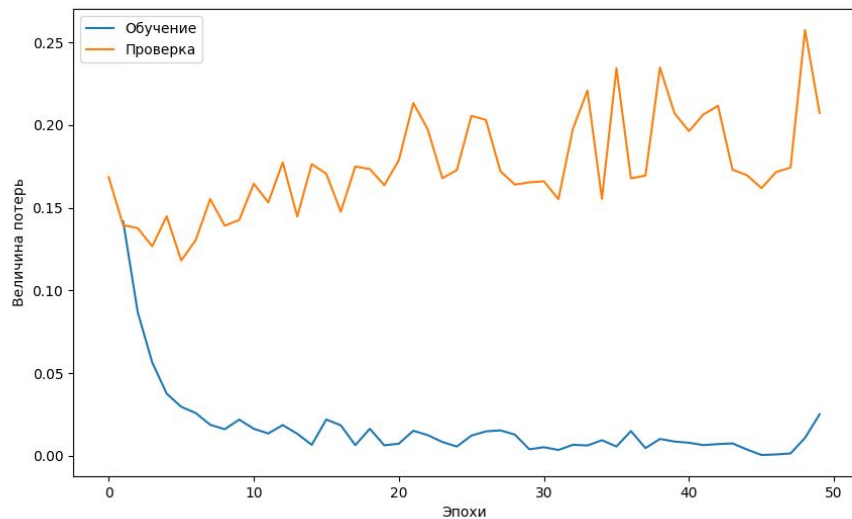
Описание эксперимента

Для анализа влияния L2-регуляризации и Dropout на процесс обучения нейронной сети мы рассмотрим классический пример задачи машинного обучения: **классификация изображений MNIST**.



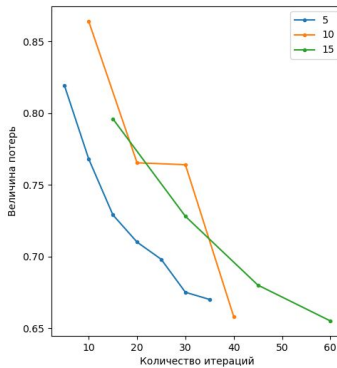
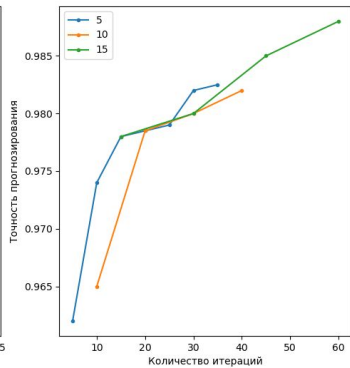
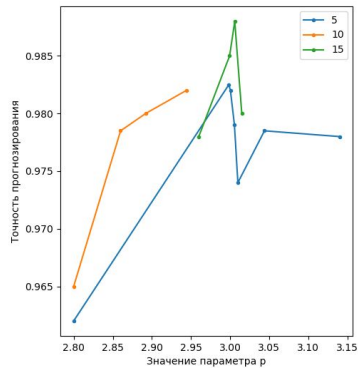
Результаты

Графики результатов обучения нейросети на данных MNIST

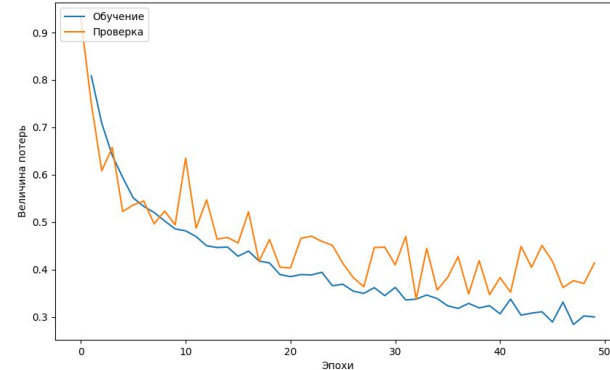


Результаты (L2)

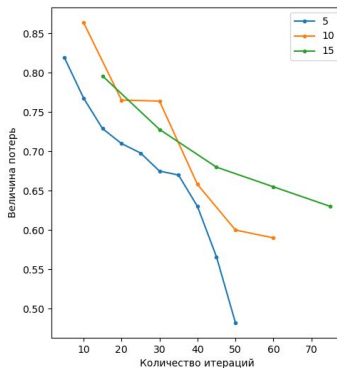
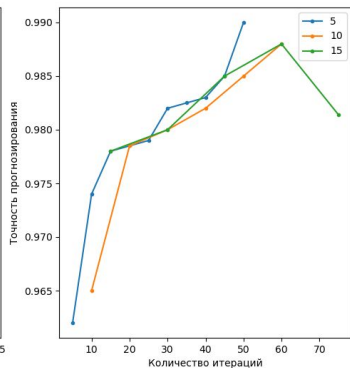
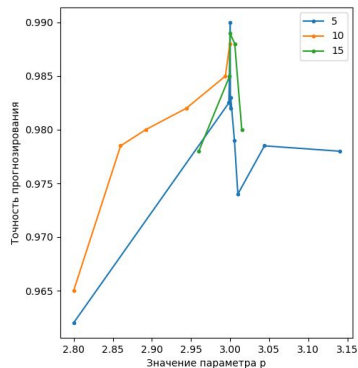
L2 регуляризация. Eps = 0.01



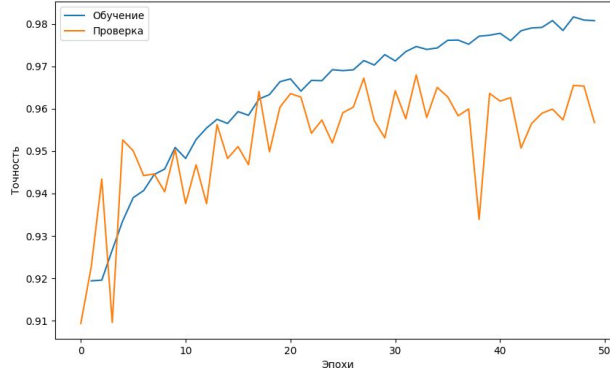
Обучение с использованием L2



L2 регуляризация. Eps = 0.001

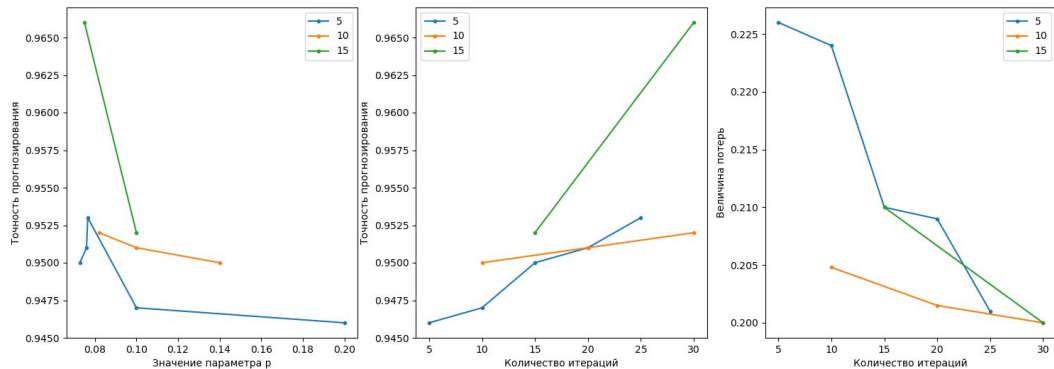


Обучение с использованием L2

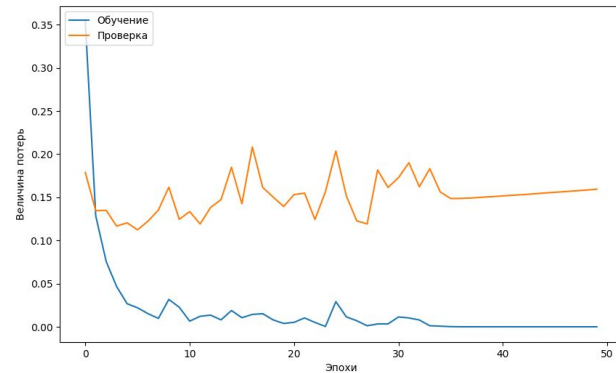


Результаты (Dropout)

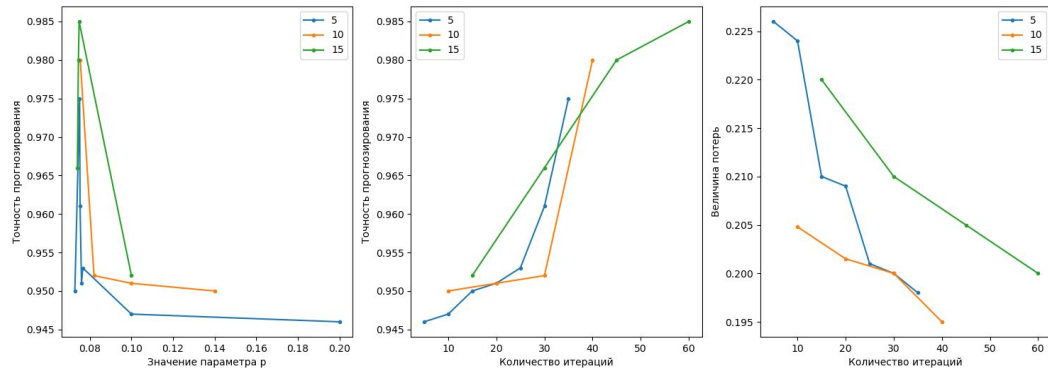
Dropout. Eps = 0.01



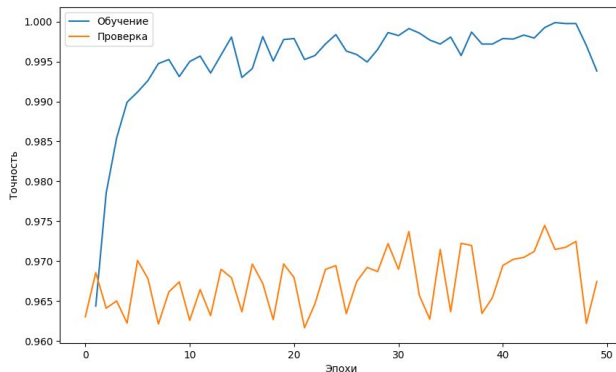
Обучение с использованием Dropout



Dropout. Eps = 0.001



Обучение с использованием Dropout



Заключение

В данной работе были подробно рассмотрены методы регуляризации L2 и Dropout. По результатам исследований можно сделать вывод о **небходимости использования таких методов для оптимизации процесса обучения** и получения более эффективных моделей.

Данная работа показала, что у методов L2 и Dropout есть **свои достоинства и недостатки**, поэтому стоит заранее продумать, какой метод регуляризации использовать при разработке собственных нейронных сетей. **Обычно используется Dropout, как стандартный метод оптимизации процесса обучения.**

Также хотим отметить, что **метод оптимизации гиперпараметров Grid Search** (его одномерный случай) **хорошо показал себя для данной задачи** и может быть использован при разработке более сложных моделей.

Список использованных источников

- [1]БРЭ. Большая Российская Энциклопедия. Нейронные сети. — URL: <https://bigenc.ru/c/neironnyeseti-e734b3> ; (дата обр. 16.04.2024).
- [2]Шолле Ф. Глубокое обучение на Python. — Санкт-Петербург : изд-во "Питер", 2023.
- [3]Liu M., Cheng G. Early Stopping for Nonparametric Testing. — 2018. — URL: <https://arxiv.org/pdf/1805.09950> ; (дата обр. 01.05.2024).
- [4]Тихонов А. Н. О методах регуляризации задач оптимального управления // Доклады Академии наук СССР. — 1965. — т. 162, № 4. — с. 3—4.
- [5]Regularization methods / A. N. Tikhonov [и др.] // Numerical Methods for the Solution of Ill-Posed Problems. — Dordrecht : Springer Netherlands, 1995. — с. 7—63.
- [6]Srivastava N., Geoffrey Hinton Alex Krizhevsky I. S., Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting // Journal of Machine Learning Research. — 2014. — т. 15, № 56. — с. 1929—1958.
- [7]Википедия. MNIST (база данных). — URL: https://en.wikipedia.org/wiki/MNIST_database ; (дата обр. 06.05.2024).