

УДК 004.896

## ОПТИМИЗАЦИЯ ФУНКЦИЙ АКТИВАЦИИ ДЛЯ ПОСТРОЕНИЯ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ ПРОГНОЗА ПОГАШЕНИЯ КРЕДИТНЫХ ОБЯЗАТЕЛЬСТВ

Май Д.Т.Т. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – Осипов Н.А., доцент, кандидат технических наук  
(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

**Аннотация.** В докладе представлены нелинейные функции активации, которые часто используются в скрытых слоях нейронных сетей. Кроме того, в докладе изучена эффективность различных структур нейронных сетей с разными функциями активации в решении задачи прогноза погашения кредитных обязательств.

**Введение.** За последние несколько лет количество плохих кредитов и мошенничества значительно увеличилось, и поэтому крайне важно, чтобы банки и финансовые учреждения использовали мощные механизмы для прогнозирования доходности ссуд. Благодаря беспрецедентному росту как банковских, так и цифровых платежей, банки и финансовые учреждения становятся архивами больших объемов разнообразных данных. Повышенная доступность данных может привести к принятию более разумных решений. Модели машинного обучения используют данные, чтобы найти правила, лежащие в основе проблемы. Можно применить теории машинного обучения как инструмента для преобразования информации в знания с целью прогнозирования будущих событий, именно прогнозирования погашения новых кредитных обязательств.

**Основная часть.** Нейронная сеть состоит из отдельных нейронов, называемых Perceptrons. Perceptron будет принимать один или несколько входов  $x$  и выводить один выход. На входы влияют их соответствующие весовые параметры  $w$ , а выходы представлены нелинейной функцией активации  $f(s)$  следующим образом:  $f(s) = f(w^T x)$ .

Сигмоидальная функция имеет вид  $f(s) = \frac{1}{1+e^{-s}}$  ( $s = w^T x$ ). Если вход большой, функция даст выход, близкий к 1. Для маленького входа (очень отрицательного) функция даст выход, близкий к 0. У него есть один существенный недостаток: Когда вход имеет большое абсолютное значение (очень отрицательное или очень положительное), градиент этой функции будет очень близок к 0. Это означает, что веса, соответствующие рассматриваемому нейрону, почти не будут обновляться.

Функция Tanh с видом  $f(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}}$  ( $s = w^T x$ ) имеет тот же недостаток, что имеет очень маленький градиент для входных данных с большими абсолютными значениями.

Функция Relu имеет математическую формулу  $f(s) = \max(0, s)$  – очень просто. Relu вычисляется почти мгновенно, и его градиент также вычисляется чрезвычайно быстро с градиентом 1, если входной сигнал больше 0, и нулевым, если входной сигнал меньше нуля.

**Выводы.** У функций активации Sigmoid и Tanh есть проблема исчезающего градиента - когда вход имеет большое абсолютное значение, градиент этих функций будет очень близок к 0, это означает, что веса, соответствующие рассматриваемому нейрону, почти не будут обновляться. Эта проблема приводит к тому, что модель мало обучится и имеет низкую эффективность, а функция активации Relu с градиентом 1, если входной сигнал больше 0, и нулевым, если входной сигнал меньше нуля, может решить эту проблему и имеет более высокую эффективность.