МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
ГОМЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ П. О. СУХОГО

Факультет автоматизированных и информационных систем
Кафедра «Информационные технологии»
Специальность 6-05-0611-01 Информационные системы и технологии (В интеллектуальном анализе данных и обработке информации)

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к курсовому проекту
по дисциплине «Введение в нейронные сети»

на тему: «Прогнозирование энергии от ветряной электростанции с использованием многослойного перцептрона»

Выполнил: студент гр. ИТД-31
Чайдаков И. М.
Руководитель:

Дата проверки: _____
Дата допуска к защите: _____
Дата защиты: _____
Оценка работы: _____

Подписи членов комиссии
по защите курсового проекта: _____

Гомель 2025

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Возрастающая доля возобновляемых источников энергии в мировом энергобалансе обуславливает необходимость решения задач точного прогнозирования выработки электроэнергии. Ветряная энергетика, являясь одним из наиболее динамично развивающихся направлений, сталкивается с фундаментальной проблемой — изменчивостью и непредсказуемостью ветровых условий. Точный прогноз выработки электроэнергии ветряными электростанциями становится критически важным для обеспечения стабильности энергосистемы, оптимизации диспетчерского управления и повышения экономической эффективности.

В рамках данного проекта используется Wind Turbine Scada Dataset, содержащий исторические данные работы ветряных турбин. Датасет включает временные ряды технологических параметров с фиксированным интервалом измерения: скорость и направление ветра, температура окружающей среды, параметры работы оборудования и фактическая выработка электроэнергии. Эти данные представляют собой ценный источник информации для построения моделей прогнозирования, позволяя выявлять сложные нелинейные зависимости между метеорологическими условиями и энергетической отдачей турбин.

Для решения задачи прогнозирования часовой выработки электроэнергии предлагается использование многослойного перцептрона — нейронной сети прямого распространения. Выбор этой архитектуры обусловлен ее способностью аппроксимировать любые непрерывные функции и выявлять сложные нелинейные закономерности в данных. Многослойный перцептрон преодолевает ограничения линейных моделей и эффективно обучается на исторических данных с помощью алгоритма обратного распространения ошибки.

Целью проекта является разработка модели прогнозирования суммарной часовой выработки электроэнергии ветряной электростанции на основе многослойного перцептрона. Ожидается, что модель позволит значительно повысить точность прогнозов по сравнению с традиционными методами, что внесет вклад в решение актуальных задач управления энергосистемами с высокой долей возобновляемой генерации.

1 ПОСТАНОВКА ЗАДАЧИ И ИСХОДНЫЕ ДАННЫЕ

1.1. Описание данных и постановка задачи

В качестве исходных данных для решения задачи прогнозирования используется Wind Turbine Scada Dataset: типичный набор данных *SCADA*-системы (Supervisory Control and Data Acquisition) ветряной электростанции. Данный датасет содержит временные ряды технологических параметров работы ветряных турбин с фиксированным интервалом измерения (обычно 10 минут).

Основные параметры, содержащиеся в датасете:

– Скорость ветра (м/с) на различных высотах;

– Направление ветра (градусы);

– Температура окружающей среды (°C);

– Температура компонентов турбины;

– Активная мощность (кВт);

– Состояние оборудования и коды ошибок;

– Угол поворота лопастей;

– Обороты ротора.

Как отмечает Хайкин, «знания представляются в самой структуре нейронной сети с помощью ее состояния активации». В нашем случае исходные данные *SCADA* системы представляют собой ту самую информацию из окружающей среды, которая используется для обучения сети.

Задача формулируется следующим образом: на основе исторических данных *SCADA*-системы за предыдущие периоды построить модель, способную прогнозировать суммарную часовую выработку электроэнергии ветряной электростанцией.

Формально: временной ряд параметров турбин $\{\mathbf{x}_t\}_{t=1}^{T}$, где x_t — вектор измерений в момент времени t_i, необходимо научиться предсказывать целевую переменную (1.1):

$$y_{t+1} = \sum_{\tau=t+1}^{t+1} P_\tau \qquad (1.1)$$

где P_T — мгновенная мощность в момент времени t, а y_{t+1} — суммарная выработка за следующий час.

1.2. Физическая модель выработки мощности ветротурбиной

Входные данные, используемые для обучения модели, представляют собой измерения SCADA-системы ветряной турбины. Для понимания закономерностей в этих данных необходимо кратко описать физическую природу формирования активной мощности. Поток воздуха плотностью ρ, проходящий через площадь ометания лопастей A, обладает кинетической энергией, пропорциональной v^2. Соответственно, мощность воздушного потока выражается как (1.2):

$$P_{\text{wind}} = \frac{1}{2}\rho A v^3 \tag{1.2}$$

где v — скорость ветра. Турбина преобразует лишь часть этой энергии, и фактическая электрическая мощность описывается соотношением (1.3):

$$P_{\text{turbine}} = C_p(\lambda, \beta) \cdot \frac{1}{2}\rho A v^3 \tag{1.3}$$

где $C_p(\lambda,\beta)$ — коэффициент использования энергии ветра, зависящий от быстродействия и аэродинамических параметров лопастей. Несмотря на сложность аэродинамической модели, зависимость мощности от скорости ветра в широком диапазоне значений хорошо аппроксимируется типичной кривой мощности (*power curve*), имеющей следующие характерные участки:

– область включения: при малой скорости ветра ($v < v_{cut\text{-}in}$) мощность равна нулю;

– рабочий диапазон: в промежутке между $v_{cut\text{-}in}$ и v_{rated} мощность возрастает приблизительно по кубическому закону;

– номинальный режим: при скорости $v_{rated} \leq v < v_{cut\text{-}out}$ мощность стабилизируется на уровне P_{rated};

– отключение: при чрезмерно высокой скорости ветра ($v \geq v_{cut\text{-}out}$) турбина останавливается.

Таким образом, при скорости выше скорости включения $v_{cut\text{-}in}$ значения активной мощности не могут быть строго нулевыми при нормальной работе установки. Ниже, приведена приблизительная модель кривой мощности ветротурбины (Рисунок 1.1).

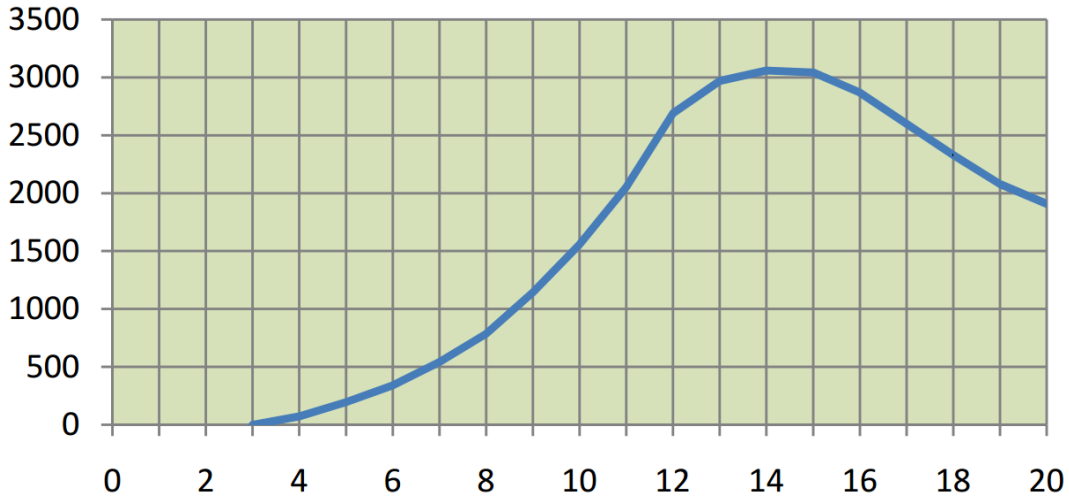

Рисунок 1.1 — Модель кривой мощности ветротурбины.

1.3. Задача регрессии и основные понятия

Задача регрессии представляет собой фундаментальный тип задач машинного обучения с учителем, целью которой является прогнозирование непрерывной числовой величины на основе входных признаков. В контексте нашей задачи, речь идет о предсказании часовой выработки электроэнергии ветряной электростанцией на основе метеорологических данных и параметров работы оборудования.

Как отмечает Хайкин, «знания о мире включают два типа информации: известное состояние окружающего мира, представленное имеющимися в наличии достоверными фактами, и наблюдения за окружающим миром, полученные с помощью сенсоров». В нашем случае это означает, что мы располагаем историческими данными *SCADA* системы о выработке электроэнергии и соответствующих технологических параметрах.

Формально задача регрессии может быть описана следующим образом: набор обучающих данных $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$, где x_i — вектор признаков (скорость ветра, направление, температура, параметры оборудования и др.), а d_i — соответствующее значение целевой переменной (суммарная часовая выработка

электроэнергии), требуется найти функцию *f(x)*, минимизирующую среднеквадратическую ошибку (1.3):

$$E = \frac{1}{N} \sum_{i=1}^{N} \left(d_i - f\left(\mathbf{x}_i \right) \right)^2 \tag{1.3}$$

1.4. Основные математические преобразования данных

В процессе подготовки данных и построения модели используются несколько стандартных математических преобразований, обеспечивающих корректное обучение искусственной нейронной сети. Ниже приведены основные из них.

1.4.1. *Min–max* нормализация

Для приведения числовых признаков к единому масштабу используется нормализация по формуле (1.4):

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1.4}$$

Где x — исходное значение признака, x_{min}, x_{max} — минимальное и максимальное значения в выборке. Это преобразование уменьшает резкие различия в масштабах входов и способствует устойчивой работе метода градиентного спуска.

1.4.2. Преобразование циклических признаков

Циклические величины (например, направление ветра в градусах) обладают свойством периодичности, что требует специального кодирования. Для этого используется преобразование в две ортогональные компоненты (1.5):

$$x_{\sin} = \sin\left(\frac{2\pi x}{T} \right), \quad x_{\cos} = \cos\left(\frac{2\pi x}{T} \right) \tag{1.5}$$

Где T — период (например, T=360° для угловой величины). Такое кодирование устраняет разрыв в точках перехода 0/360°.

1.4.3. Среднеквадратическая ошибка

В задаче регрессии функцией ошибки (функцией потерь) является среднеквадратическая ошибка (1.6):

$$E = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{1.6}$$

Этот критерий широко применяется в алгоритмах обучения нейронных сетей благодаря дифференцируемости и удобству вычисления градиента.

1.4.4. Обратная нормализация целевой переменной

При необходимости восстановить физическое значение величины после предсказания используется обратное преобразование (1.7):

$$y = y_{\text{norm}} \ \left(y_{\max} - y_{\min} \right) + y_{\min} \tag{1.7}$$

1.4.5. Зависимость мощности ветрового потока от скорости ветра

При анализе *SCADA*-данных используется общая формула кинетической мощности ветрового потока (1.8):

$$P_{\text{wind}} \ = \frac{1}{2} \rho A v^3 \tag{1.8}$$

А также типичная форма функциональной зависимости мощности турбины от скорости ветра (power curve), которая в упрощенном виде описывается следующей кусочно-непрерывной функцией (1.9):

$$P(v) = \begin{cases} 0, & v < v_{\text{cut-in}} \\ f(v), & v_{\text{cut-in}} \ \leq v < v_{\text{rated}} \ , \\ P_{\text{rated}} \ , & v_{\text{rated}} \ \leq v < v_{\text{cut-out}} \ , \\ 0, & v \geq v_{\text{cut-out}} \ . \end{cases} \tag{1.9}$$

Эта модель используется далее при обосновании удаления аномальных записей.

1.5. Перцептрон: от бинарного классификатора к универсальному аппроксиматору

Перцептрон, впервые описанный Розенблаттом, представляет собой фундаментальную модель искусственного нейрона. Согласно Хайкину, «нейронная сеть — это громадный распределенный параллельный процессор, состоящий из элементарных единиц обработки информации». Базовая модель нейрона включает три ключевых компонента:

– Синаптические веса w_{kj}, определяющие силу связи;

– Сумматор для линейной комбинации входных сигналов;

– Функцию активации $\varphi(\cdot)$, ограничивающую амплитуду выходного сигнала.

Математически это описывается парой уравнений (1.10 и 1.11):

$$v_k = \sum_{j=1}^{m} w_{kj} x_j + b_k \tag{1.10}$$

$$y_k = \varphi\left(v_k\right) \tag{1.11}$$

Однослойный перцептрон обладал фундаментальным ограничением — он мог решать только линейно разделимые задачи. Как отмечает Хайкин, «в 1969 году вышла книга Минского и Пейперта, в которой математически строго обоснованы фундаментальные ограничения однослойного перцептрона».

Классическим примером является задача «исключающего ИЛИ» (*XOR*), для которой невозможно найти единственную разделяющую гиперплоскость. Это ограничение привело к разработке многослойных архитектур, где «функция последних (скрытых нейронов) заключается в посредничестве между внешним входным сигналом и выходом нейронной сети».

1.6. Многослойный перцептрон для решения задачи регрессии

Многослойный перцептрон (МЛП) преодолевает ограничения однослойной архитектуры за счет введения скрытых слоев. Как подчеркивает Хайкин, «добавляя один или несколько скрытых слоев, мы можем выделить статистики высокого порядка».

Для задачи регрессии прогнозирования выработки электроэнергии ключевыми особенностями архитектуры являются:

– Входной слой: количество нейронов соответствует размерности вектора признаков из *SCADA* системы;

– Скрытые слои: используют нелинейные функции активации (сигмоида, гиперболический тангенс) для выявления сложных зависимостей;

– Выходной слой: содержит один нейрон с линейной функцией активации для прогнозирования непрерывного значения выработки.

Согласно теореме об универсальной аппроксимации, МЛП с хотя бы одним скрытым слоем и нелинейной функцией активации может аппроксимировать любую непрерывную функцию с любой точностью. Это делает его идеальным инструментом для решения задач нелинейной регрессии в энергетике.

1.7. Механика обратного распространения ошибки с использованием функции активации ReLU

Алгоритм обратного распространения ошибки, подробно рассмотренный С. Хайкиным, представляет собой фундаментальный механизм настройки параметров многослойных нейронных сетей. Его ключевая идея заключается в том, что ошибка прогноза на выходе сети распространяется назад — от последнего слоя к первому — и используется для корректировки весов. Это позволяет сети «учиться» прогнозировать выработку электрической энергии по данным *SCADA*-системы ветротурбины.

Процесс обучения многослойного перцептрона включает две фазы:

– Прямой проход. Вычисление выходного значения сети по входным признакам $x=(x_1,x_2,x_3)$.

– Обратный проход. Вычисление локальных градиентов и корректировка весов.

1.7.1. Используемая функция активации: *ReLU*

В разработанном программном комплексе в скрытых слоях применяется функция активации *ReLU* (1.12):

$$\varphi(v) = \max(0, v) \tag{1.12}$$

её производная имеет вид (1.13):

$$\varphi'(v) = \begin{cases} 1, & v > 0 \\ 0, & v \leq 0 \end{cases} \tag{1.13}$$

что делает *ReLU* простой в вычислении и устойчивой в обучении — особенно на больших выборках *SCADA*-данных.

1.7.2. Вычисление ошибки

Для задачи регрессии с линейным выходным нейроном используется стандартная ошибка (1.14):

$$e = y - \hat{y} \tag{1.14}$$

где
y — истинное значение активной мощности,
ŷ — предсказание сети.

Среднеквадратичная ошибка (1.15):

$$E = \frac{1}{2} e^2 \tag{1.15}$$

Поскольку выходной нейрон линейный, его производная равна 1 (1.16):

$$\varphi'\left(v_{\text{out}}\right) = 1 \tag{1.16}$$

Поэтому дельта выходного слоя вычисляется как (1.17):

$$\delta_{\text{out}} = e \tag{1.17}$$

1.7.3. Дельта-правило для скрытых слоев с *ReLU*

На скрытых слоях локальный градиент рассчитывается как (1.18):

$$\delta_j = \varphi'(v_j) \sum_k \delta_k w_{jk} \tag{1.18}$$

где
w_{jk} — вес между нейронами слоёв j и k,
δ_k — ошибка следующего слоя,
$\varphi'(v_j)$ — производная ReLU из формулы (1.12).

Поскольку *ReLU* обнуляет производную для отрицательных значений, нейроны со значением $v_j \leq 0$ перестают распространять ошибку назад. Это делает обучение устойчивым и устраняет проблему затухающих градиентов, характерную для сигмоиды.

1.7.4. Корректировка весов

После вычисления локальных градиентов выполняется обновление весов и смещений (1.19 и 1.20):

$$\Delta w_{ij} = -\eta \delta_j x_i \tag{1.19}$$

$$\Delta b_j = -\eta \delta_j \tag{1.20}$$

где
η — скорость обучения,
x_i — входы текущего слоя.

Эта формулировка полностью соответствует классическому алгоритму обратного распространения Хайкина, с учётом выбранной функции активации.

1.7.5. Пример вычисления локального градиента (*ReLU*)

Пусть на скрытом нейроне:

Так как $v_j > 0$, производная *ReLU*:

$$\varphi'(v_j) = 1$$

Пусть суммарный вклад ошибок следующего слоя:

$$\sum_k \delta_k w_{jk} = 0.5$$

Тогда локальный градиент скрытого нейрона:

$$\delta_j = 1 \cdot 0.5 = 0.5$$

Эта дельта используется далее в формулах (1.19–1.20) для обновления весов и смещений предыдущего слоя.

1.7.6. Преимущества *ReLU* для задачи прогнозирования *SCADA*-данных

– отсутствие насыщения градиентов в положительной области;

– разреженная активация — многие нейроны автоматически «выключаются», снижая переобучение;

– высокая вычислительная эффективность;

– хорошая сходимость при большом объеме данных (более 40 тыс. измерений в *T1.csv*).

Эти свойства делают *ReLU* предпочтительным выбором для практической реализации модели в условиях реальных эксплуатационных данных ветротурбин.

2 РАЗРАБОТКА И РЕАЛИЗАЦИЯ ПРОГРАММНОГО КОМПЛЕКСА

2.1. Архитектура программного комплекса

Программный комплекс разрабатывался с учётом требований модульности, расширяемости и прозрачности внутренней логики. Структура проекта организована таким образом, чтобы каждый модуль выполнял строго определённую роль, не нарушая принцип единственной ответственности. Основной язык реализации — *Python*, а в качестве базовых библиотек для работы с числовыми данными применяются *NumPy* и *Pandas*, обеспечивающие удобные структуры данных и эффективные векторизованные операции.

Общая структура проекта представлена следующим образом (Рисунок 2.1):

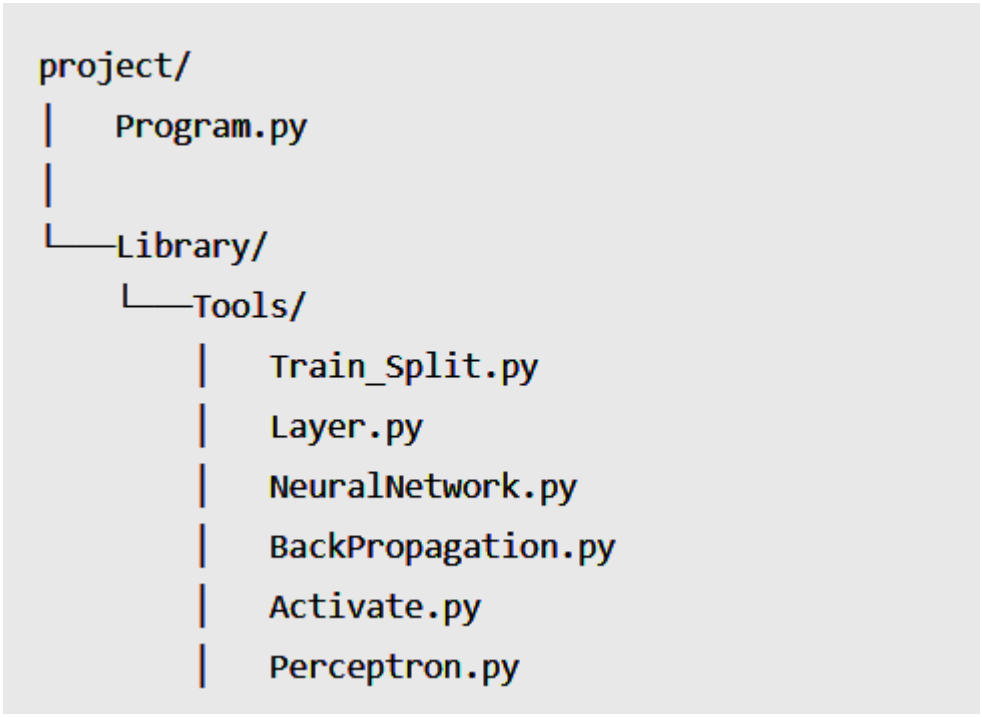

Рисунок 2.1 — Общая структура проекта.

Функции модулей следующие:

– *Train_Split.* предобработка данных, нормализация и формирование обучающей и тестовой выборок.

– *Layer*. Реализация одного полносвязного слоя многослойного перцептрона.

– *NeuralNetwork*. Реализация общей архитектуры *MLP* и процедуры обучения.

– *BackPropagation*. Реализация алгоритма обратного распространения ошибки согласно теории Хайкина.

– *Activate*. Функции активации и их производные.

– *Perceptron*. Базовая структура нейрона (используется как элемент слоя).

– *Program*. Управляющий сценарий, инициирующий обработку данных, обучение и тестирование.

Такое разделение облегчает сопровождение кода, позволяет локализовать изменения и упрощает последующее развитие проекта (см. Приложение В).

2.2. Предобработка и анализ исходных данных

Этот этап является критически важным при построении любой модели машинного обучения. Исходный датасет *Wind Turbine SCADA Dataset* (см. Приложение А) содержит реальные измерения, полученные с управляющей системы ветротурбины. Данные имеют временную структуру с шагом в 10 минут, что требует особого обращения.

Процесс предобработки включает несколько последовательных этапов.

2.2.1. Фильтрация и очистка данных

В датасете встречаются физически невозможные или аномальные значения. На предварительном анализе были выявлены следующие случаи:

– Отрицательная активная мощность (*LV ActivePower* < 0)**.** Такие записи не имеют физического смысла и исключаются из выборки.

– Ситуации, когда скорость ветра > 2 м/с, а активная мощность равна нулю. Согласно теоретическому описанию в первой главе (1.2), при скорости свыше $v_{cut\text{-}in}$ мощности не может быть строго нулевой. Значит, это вероятный сбой датчика или режим обслуживания.

Удаление подобных выбросов предотвращает искажение распределения данных и снижает шум, что положительно влияет на сходимость алгоритма.

2.2.2. Нормализация признаков

В рамках задачи регрессии важно, чтобы все входные признаки имели сопоставимый масштаб. Для скорости ветра используется *min–max* нормализация (1.4). Это соответствует рекомендациям Хайкина при использовании сигмоидальных или ReLU-подобных активаций.

2.2.3. Кодирование циклических признаков

Направление ветра — циклическая величина. Использование её в исходном виде (0–360°) приводит к ложной дискретности. Поэтому применяется преобразование (1.5). Этот приём повсеместно используется при работе с циклическими признаками и соответствует рекомендациям по кодированию периодических параметров.

Полученные синусная и косинусная компоненты также нормализуются в диапазон [0, 1].

2.2.4. Финальное формирование признаков

После преобразований удаляются лишние столбцы:

– *Date/Time*. Временная метка, не используемая как числовой признак;

– *Theoretical_Power_Curve*. Теоретическая модель мощности, которая не должна быть подана в сеть, чтобы не вносить утечку информации;

– *Wind Direction (°)*. Заменён двумя компонентами.

В результате получается компактный и информативный набор: *Wind Speed (m/s), Wind Direction_sin*, *Wind Direction_cos*. Целевая переменная *LV ActivePower (kW)* дополнительно нормализуется (Рисунок 2.2).

```
Первые 5 строк данных:
          Date/Time  LV ActivePower (kW)  Wind Speed (m/s)  Theoretical_Power_Curve (KWh)  Wind Direction (°)
0  01 01 2018 00:00           380.047791          5.311336                     416.328908          259.994904
1  01 01 2018 00:10           453.769196          5.672167                     519.917511          268.641113
2  01 01 2018 00:20           306.376587          5.216037                     390.900016          272.564789
3  01 01 2018 00:30           419.645905          5.659674                     516.127569          271.258087
4  01 01 2018 00:40           380.650696          5.577941                     491.702972          265.674286
Всего записей: 43283

Данные после предобработки:
   LV ActivePower (kW)  Wind Speed (m/s)  Wind Direction_sin  Wind Direction_cos
0           380.047791          0.210717            0.007604            0.413132
1           453.769196          0.225032            0.000141            0.488143
2           306.376587          0.206936            0.000501            0.522375
3           419.645905          0.224537            0.000121            0.510978
4           380.650696          0.221294            0.001424            0.462287
Форма данных после предобработки: (43283, 4)

Формы перед батчингом:
X: (43283, 3), y: (43283,)
Колонки X: ['Wind Speed (m/s)', 'Wind Direction_sin', 'Wind Direction_cos']

Группировка данных:
Всего записей: 43283
Количество полных батчей: 7213
Используется записей: 43278
Отброшено записей: 5
```

Рисунок 2.2 — Данные до и после нормализации.

2.2.5. Использование временной структуры данных

Так как исходные данные представлены временными последовательностями, нельзя перемешивать строки напрямую — это может нарушить временную корреляцию. В работе применяется блочный подход:

– Данные делятся на блоки по 6 строк (60 минут);

– Блоки перемешиваются как единые элементы.

При этом обучение сети происходит по отдельным строкам — батчи применяются только как инструмент корректного перемешивания.

Такой подход сохраняет структуру данных и исключает взаимную утечку информации между тестовой и обучающей выборками.

2.3. Архитектура многослойного перцептрона

В соответствии с теоретическими положениями главы 1 используется многослойный перцептрон (*MLP*), являющийся универсальным аппроксиматором непрерывных функций (1.10).

В работе применяется архитектура [3, 16, 8, 4, 1], где:

– 3 входных признака;

– 3 скрытых слоя с убывающим числом нейронов;

– 1 выходной линейный нейрон.

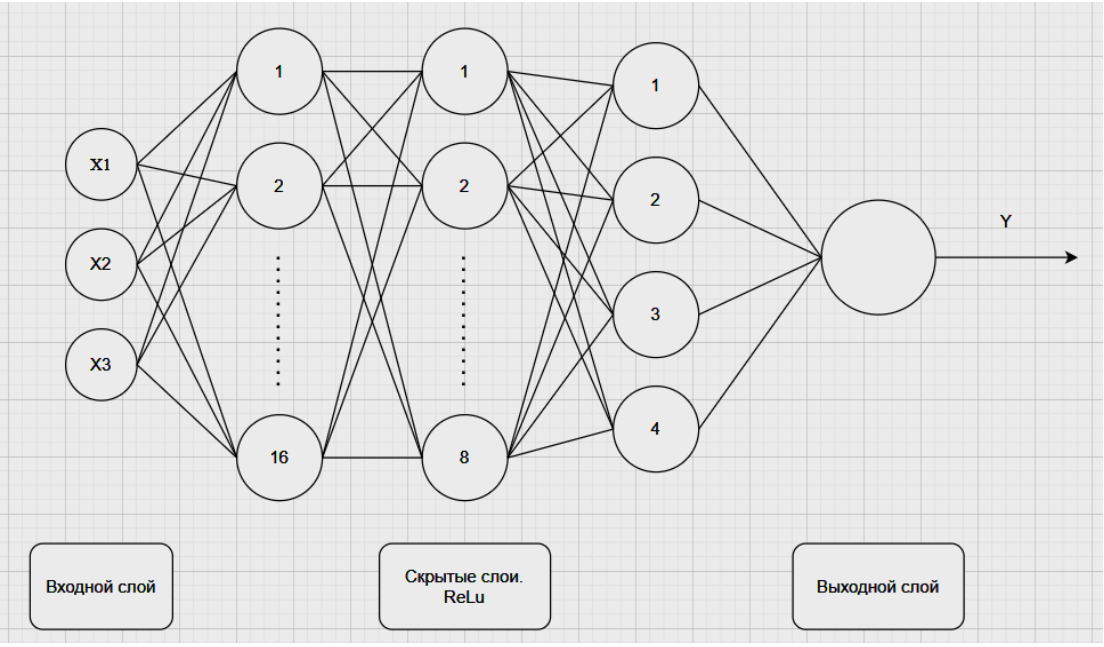

Рисунок 2.3 — Архитектура сети.

2.3.1. Обоснование выбора архитектуры

Постепенная компрессия слоев (16 → 8 → 4) позволяет слою приближать всё более сложные нелинейные проявления зависимости между скоростью, направлением ветра и выходной мощностью.

Наличие нескольких скрытых слоев позволяет сети моделировать нелинейные зависимости высокого порядка, что соответствует теоретическим выводам о мощности глубоких сетей по Хайкину.

Выходной слой является линейным, что соответствует постановке задачи регрессии (1.3).

2.4. Реализация слоя нейронной сети

Класс Layer реализует один полносвязный слой, состоящий из нескольких перцептронов. Основные элементы:

– *Weights*. Матрица весов;

– *Biases* — вектор смещений;

– *Activate_function* — функция нелинейности;

– *Forward*() — прямое распространение.

Реализация учитывает сохранение промежуточных значений (*z*, *output*, *last_input*), необходимых для выполнения обратного распространения ошибки (1.17).

2.5. Функции активации

В разработанном программном комплексе функции активации вынесены в отдельный модуль *Activate.py*, что обеспечивает независимость их реализации от остальной архитектуры сети. Функции активации являются ключевыми элементами многослойного перцептрона, поскольку именно они создают нелинейность, позволяющую нейронной сети аппроксимировать сложные функциональные зависимости, что согласуется с теоретическими выводами главы 1.

Основной функцией активации в скрытых слоях является *ReLU* (*Rectified Linear Unit*). Формальное определение функции *ReLU* и её производной приведено в разделе 1.7.1, где изложена соответствующая математическая база. Использование *ReLU* обусловлено её преимуществами при обучении глубоких моделей: функция не насыщается в положительной области, ее производная

принимает простую форму, а вычислительная сложность минимальна. Как показано в разделе 1.7, эти свойства способствуют устойчивости градиента при обратном распространении ошибки и улучшению сходимости градиентного спуска по сравнению с сигмоидой, рассмотренной в классических моделях.

Применение *ReLU* в скрытых слоях позволяет эффективно реализовать шаги расчета локальных градиентов и обновления весов, описанные в формулах (1.18-1.20). Поскольку производная *ReLU* равна нулю в отрицательной области входного сигнала, происходит естественная разреженность активации — часть нейронов «выключается», что снижает риск переобучения и соответствует современным практикам построения глубоких сетей.

Для обеспечения гибкости и возможности расширения в модуле *Activate.py* также реализованы дополнительные функции активации:

– *sigmoid*, использовавшаяся в ранних версиях архитектуры и рассмотренная теоретически в пункте 1.5;
– *tanh*, пригодная для задач, требующих симметричного диапазона выходов;
– соответствующие производные, используемые при вычислении локальных градиентов в алгоритме обратного распространения.

Хотя альтернативные функции активации оставлены в составе проекта для возможных экспериментов, основная реализация многослойного перцептрона использует именно *ReLU*, поскольку она обеспечивает наилучшее сочетание стабильности, скорости обучения и качества аппроксимации данных *SCADA*-системы.

2.6. Реализация алгоритма обучения

Алгоритм обучения многослойного перцептрона является ключевым компонентом программного комплекса и полностью основан на подходе обратного распространения ошибки, изложенном в теоретической части (раздел 1.7). Реализация обучения в программной среде следует общепринятой схеме: прямой проход по слоям сети, вычисление ошибки, обратный проход и обновление параметров. Такой порядок обеспечивает корректную минимизацию функции ошибки и последовательное улучшение качества прогноза.

В программной реализации обучение выполняется методом градиентного спуска с фиксированной скоростью обучения. Каждая итерация градиентного спуска (эпоха) включает следующие шаги:

– Прямое распространение сигнала. Входные данные из обучающей выборки последовательно проходят через слои сети, где к каждому слою применяется соответствующая функция активации. В скрытых слоях используется функция *ReLU*, свойства которой позволяют эффективно и стабильно вычислять выходы даже при большом количестве слоёв. Вычисленные значения активаций сохраняются во внутреннем состоянии каждого слоя, что необходимо для последующего вычисления градиентов.

– Вычисление ошибки на выходном слое. Поскольку задача носит регрессионный характер, ошибка определяется как разность между фактическим значением активной мощности и предсказанием сети. Среднеквадратичное отклонение используется как количественная метрика точности, формально определенная в разделе 1.6. Вычисленная ошибка служит отправной точкой для формирования локального градиента выходного слоя (см. формулу (1.15)).

– Обратное распространение градиента. В соответствии с формальной схемой, приведённой в разделе 1.7, сначала вычисляется дельта выходного слоя, а затем — дельты скрытых слоев. Для скрытых слоев используется выражение (1.16), в котором учитывается производная функции активации *ReLU*. Благодаря простоте её производной (см. формулу (1.12)), процесс вычисления локальных градиентов становится особенно эффективным: нулевые значения производной приводят к «отключению» части нейронов, что уменьшает шум градиента и делает алгоритм устойчивым на больших выборках.

– Обновление весов и смещений. На последнем шаге производится корректировка параметров сети согласно выражениям (1.17-1.18). Все обновления выполняются в направлении уменьшения функции ошибки. Параметры корректируются векторизованным образом, что обеспечивает высокую вычислительную эффективность и позволяет работать с большими объемами данных без существенных затрат времени.

Важной особенностью реализации является использование векторных вычислений библиотеки NumPy, что минимизирует количество циклов и повышает производительность на несколько порядков по сравнению с наивной реализацией. Благодаря этому обучение сети на выборке, содержащей несколько тысяч наблюдений, выполняется достаточно быстро, что позволяет экспериментировать с гиперпараметрами и архитектурой.

Реализация алгоритма обучения предусматривает ведение журнала ошибок (при включенном режиме *graph=True*). Это позволяет отслеживать динамику сходимости модели, выявлять возможные проблемы переобучения и корректировать параметры обучения. В частности, в процессе тестирования было установлено, что при скорости обучения 0.1 и количестве эпох 2000 наблюдается устойчивая сходимость модели, а ошибка на тестовой выборке остается в приемлемых пределах.

Дополнительным преимуществом реализованной схемы обучения является модульность. Алгоритм обратного распространения расположен в отдельном файле *BackPropagation.py*, что позволяет независимо изменять его логику, подключать различные оптимизаторы (например, *Adam* или *RMSProp*), добавлять регуляризацию или адаптировать алгоритм под другие типы архитектур. Такой подход обеспечивает масштабируемость и расширяемость системы, что особенно важно для дальнейших исследований и модификаций.

Таким образом, реализация обучения в программном комплексе полностью соответствует теоретическим положениям, изложенным в первой главе, и обеспечивает корректное вычисление градиентов, стабильную сходимость и высокую гибкость в последующем развитии системы.

2.7. Процесс тестирования и анализ результатов

После завершения обучения осуществляется тестирование на отложенной части выборки. Процесс тестирования включает выполнение прямого распространения сигнала через все слои модели без обновления параметров. Полученные предсказания нормализованной мощности преобразуются обратно в исходные значения в соответствии со схемой линейного восстановления, приведенной в разделе 1.4.

Качество работы модели оценивается с помощью метрики среднеквадратичной ошибки (*MSE*), определение которой приведено в пункте 1.6 теоретической части. Поскольку задача носит регрессивный характер, *MSE* является наиболее информативным показателем средней величины ошибки предсказания.