MLP-калькулятор

Рекомендации по параметрам обучения

Многослойный перцептрон – руководство по гиперпараметрам

Русский

Вольфганг Бёме

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

2 ноября 2026 г.

Документация на этом сайте разработана при поддержке ИИ Claude (Anthropic).

Содержание

1 Основные концепции **2**

2 Примеры наборов данных в калькуляторе v20 **2**

3 Кросс-курсы (UZS): 4 входа → EUR/UZS **2**

4 Валютные данные (UZS): 4 входа → EUR **3**

5 Логические вентили: XOR, AND, OR **5**

6 Классификация по кругу (200 выборок) **5**

7 Простая регрессия – синусоида (50 выборок) **6**

8 Другие типичные варианты применения **6**
8.1 Классификация цветков ириса (3 класса) 7
8.2 Регрессия временных рядов (напр. температура, акции) 7
8.3 Прогнозирование цен на жильё (несколько признаков) 8
8.4 Обнаружение спама / Бинарная классификация текста 8

9 Общие практические правила **9**

Основные концепции

Обзор функций активации:

- **ReLU** (*Rectified Linear Unit*): $f(x) = \max(0, x)$ – быстрая, стабильная, стандарт для скрытых слоёв при регрессии
- **Sigmoid**: $f(x) = \frac{1}{1+e^{-x}}$ – выход в диапазоне $[0, 1]$, подходит для бинарной классификации
- **Tanh**: $f(x) = \tanh(x)$ – выход в $[-1, 1]$, хорош для симметричных данных и временных рядов
- **Linear**: $f(x) = x$ – обязателен для непрерывной регрессии в выходном слое

Функции потерь:

- **MSE** (*Mean Squared Error*): стандарт для регрессии, сильно штрафует большие отклонения
- **BCE** (*Binary Cross Entropy*): стандарт для бинарной классификации (выход с Sigmoid)

Примеры наборов данных в калькуляторе v20

Автоматическая настройка: при загрузке примера (вкладка «Данные») калькулятор применяет рекомендуемые в этом документе настройки и создаёт сеть. Если какая-либо настройка затем отклоняется от рекомендации, вкладка «Конфигурация» показывает таблицу «рекомендуется / текущее» с предупреждающими знаками; кнопка «Применить рекомендуемые настройки» восстанавливает значения. Для собственных CSV-файлов калькулятор выводит рекомендацию из данных (эмпирические правила — в разделе 9).

Воспроизводимость: по умолчанию заданы начальное значение (seed) 42 и перемешивание обучающих данных в каждой эпохе. Тот же seed и те же настройки дают точно тот же результат; так получены все числа в этом документе.

Кросс-курсы (UZS): 4 входа → EUR/UZS

Применение: USD/UZS, EUR/USD, GBP/UZS, CNY/UZS как входы — EUR/UZS как выход. 251 торговый день (03.11.2025–30.10.2026), синтетические данные со столбцом даты.

Описание данных: курсы колеблются вокруг стабильного уровня без сильного тренда. Цель следует соотношению кросс-курса

$$\text{EUR/UZS} \approx \text{EUR/USD} \times \text{USD/UZS}$$

с небольшим шумом (стандартное отклонение 0,04 %). GBP/UZS и CNY/UZS не

влияют на цель — сеть должна сама найти значимые входы. Столбец даты при загрузке автоматически игнорируется.

Цель MLP: сеть должна выучить (почти) мультипликативную зависимость по данным. Поскольку цель определяется двумя входами, возможен прогноз заметно лучше, чем по простым правилам. Пример показывает, как в бэктестинге выглядит *хорошо прогнозируемый* набор данных.

Параметр	Рекомендация (по умолчанию)
Архитектура	4 → 8–16 → 4–8 → 1 (по умолчанию: 4 → 12 → 6 → 1)
Активация скрытых слоёв	ReLU
Активация выходного слоя	**Linear**
Функция потерь	MSE
Скорость обучения	0,001 – 0,005 (0,003)
Эпохи	2000 – 5000 (2000)
Размер батча	16 – 32 (16)
Нормализация	**активна**
Разбивка Train/Test	80 / 20, хронологически (последняя 51 строка = тест)

Результат бэктестинга (настройки по умолчанию, seed 42):

Метод (тестовый период, 51 день)	MAE	MAPE	R^2
MLP	5,34	0,037 %	0,996
Значение предыдущей строки («как вчера»)	50,92	0,354 %	0,621
Последнее значение перед тестовым периодом	80,15	0,558 %	$-0{,}0005$
Среднее обучающих данных	85,88	0,597 %	$-0{,}069$

Сеть заметно лучше всех простых правил (RMSE примерно на 90 % ниже лучшего правила). Обучение и тест близки (MAPE 0,033 % и 0,037 %) — переобучения нет.

> **Экстраполяция:** в 27 из 51 тестовой строки хотя бы один вход лежит вне обучающего диапазона (USD/UZS до 14,8 % ширины диапазона ниже, GBP/UZS до 24,7 % выше). Ошибка этих строк тем не менее не больше (MAE 5,03 против 5,68): GBP/UZS не влияет на цель, а зависимость от USD/UZS сразу за границей ещё почти линейна.

Валютные данные (UZS): 4 входа → EUR

Применение: USD/UZS, CNY/UZS, GBP/UZS, CHF/UZS как входы — EUR/UZS как выход. 251 строка (файл `Währungsdaten_NN_4-1.csv`), непрерывные значения, узкие диапазоны.

Описание данных: дневные курсы четырёх валют к узбекскому суму примерно за год. Значения лежат в узких диапазонах (напр. USD: 12 394–12 624, CNY: 1 716–1 749). *Уровни* курсов отчасти движутся вместе, однако *дневные изменения* EUR и четырёх входных курсов практически не связаны (корреляция ≈ 0). Кроме того, выходные дни содержатся как трижды повторённые строки.

Цель MLP: сеть должна оценить курс EUR/UZS по четырём курсам — *задача регрессии*. Из-за отсутствия связи в дневных изменениях набор данных прежде всего *учебный пример*: бэктестинг показывает, что малая ошибка на обучении ещё не означает пригодного прогноза.

Параметр	Рекомендация (по умолчанию)
Архитектура	$4 \to 8$–$16 \to 8$–$16 \to 1$ (по умолчанию: $4 \to 16 \to 8 \to 1$)
Активация скрытых слоёв	ReLU
Активация выходного слоя	**Linear** (обязателен при непрерывном выходе)
Функция потерь	MSE
Скорость обучения	0,001 – 0,005 (0,003)
Эпохи	2000 – 5000 (3000)
Размер батча	16 – 32 (16)
Нормализация	**всегда активна** (значения в диапазоне 12000–16000)
Разбивка Train/Test	80 / 20, хронологически

Результат бэктестинга (настройки по умолчанию, seed 42):

Метод (тестовый период)	MAE	MAPE	R^2
MLP (4 входа)	75,43	0,557 %	$-4{,}24$
MLP со значением предыдущего дня как 5-м входом*	16,23	0,120 %	0,638
Значение предыдущей строки («как вчера»)	15,85	0,117 %	0,618
Последнее значение перед тестовым периодом	31,04	0,230 %	$-0{,}005$
Среднее обучающих данных	69,40	0,512 %	$-3{,}52$

* Тестовый период 50 строк вместо 51, так как при опции лага первая строка отбрасывается; правило «как вчера» там даёт MAE 15,32, MAPE 0,113 %, R^2 0,633.

Оценка: с четырьмя входами сеть хуже постоянного значения ($R^2 < 0$); калькулятор также сообщает о возможном переобучении (MAPE на обучении 0,255 %, на тесте 0,557 %). С опцией «Использовать значение цели за предыдущий день как дополнительный вход» (вкладка «Данные») сеть достигает примерно качества правила «как вчера» — больше из этих данных получить нельзя. Поэтому MAPE ниже 1 % для валютных курсов *не* является признаком качества; решающее значение имеет сравнение с простыми правилами.

Примечание: вкладка «Прогноз» показывает для каждого входа обучающий диапазон и предупреждает об экстраполяции, указывая отклонение (абсолютно и в процентах ширины диапазона). В тестовом периоде GBP/UZS в 18 строках превышает обучающий максимум до 14,5 %.

Логические вентили: XOR, AND, OR

Классические учебные примеры бинарной классификации с 2 входами и 4 выборками.

Описание данных: Таблицы истинности булевой логики. Два бинарных входа (0 или 1) и один бинарный выход. AND даёт 1 только если оба входа равны 1; OR даёт 1 если хотя бы один вход равен 1; XOR даёт 1 если ровно один вход равен 1. Набор данных охватывает все четыре возможные комбинации входов.

Цель MLP: Сеть должна научиться выполнять логическую функцию. XOR особенно значим: он *не является линейно разделимым* и не может быть решён сетью без скрытых слоёв. Пример наглядно демонстрирует, почему многослойные сети необходимы – исторически важный результат в исследованиях ИИ.

Параметр	Рекомендация
Архитектура	2 → 4–8 → 4–8 → 1 (калькулятор требует не менее 2 скрытых слоёв)
Активация скрытых слоёв	Sigmoid
Активация выходного слоя	Sigmoid
Функция потерь	BCE
Скорость обучения	0,1
Эпохи	5000 – 10000 (XOR требует больше)
Размер батча	1 – 4 (все выборки)
Нормализация	выключена (данные уже 0/1)

Примечание по XOR: XOR не является линейно разделимым – необходим как минимум 1 скрытый слой с ≥ 4 нейронами. AND и OR линейно разделимы и теоретически могут быть решены без скрытого слоя.

Классификация по кругу (200 выборок)

200 случайных точек в $[-1,1]^2$. Выход 1, если точка внутри круга радиуса $\sqrt{0{,}5}$, иначе 0.

Описание данных: 200 случайно сгенерированных точек в квадрате $[-1,1]^2$. Точка принадлежит классу 1 (внутри), если $x^2+y^2<0{,}5$, иначе классу 0 (снаружи). Граница решения – окружность, типично нелинейная структура, которую линейные классификаторы не могут освоить.

Цель MLP: Сеть должна научиться нелинейной границе решения в двумерном пространстве. Это *задача бинарной классификации*, демонстрирующая, как MLP может аппроксимировать границы произвольной формы. Хорошо подходит для экспериментов с глубиной сети и количеством нейронов.

Параметр	Рекомендация
Архитектура	2 → 8–16 → 8 → 1
Активация скрытых слоёв	ReLU
Активация выходного слоя	Sigmoid
Функция потерь	BCE
Скорость обучения	0,01
Эпохи	2000 – 3000
Размер батча	16 – 32
Нормализация	выключена (данные в $[-1, 1]$)
Разбивка Train/Test	80 / 20

Простая регрессия – синусоида (50 выборок)

$y = 0{,}5 + 0{,}4 \cdot \sin(6x) + \varepsilon$, *где* $x \in [0, 1]$ *и шум* ε.

Описание данных: 50 равномерно распределённых точек вдоль затухающей синусоиды с наложенными случайными помехами. Вход x меняется от 0 до 1, выход y колеблется от примерно 0,1 до 0,9. Кривая содержит несколько периодов и проверяет способность сети улавливать периодические структуры.

Цель MLP: Сеть должна аппроксимировать непрерывную нелинейную функцию по зашумлённым выборкам – классическая *задача регрессии*. Демонстрирует разницу между недообучением и переобучением: слишком мало нейронов – кривая не воспроизводится; слишком много – сеть запоминает шум.

Параметр	Рекомендация
Архитектура	1 → 8–16 → 8 → 1
Активация скрытых слоёв	Tanh (симметрична, хороша для синусоиды)
Активация выходного слоя	Linear
Функция потерь	MSE
Скорость обучения	0,01
Эпохи	3000 – 5000
Размер батча	8 – 16
Нормализация	выключена (значения в $[0, 1]$)

Другие типичные варианты применения

Классификация цветков ириса (3 класса)

4 ботанических признака → 3 класса (Setosa, Versicolor, Virginica). 150 выборок.

Описание данных: 150 цветков ириса с 4 признаками: длина/ширина чашелистика и длина/ширина лепестка (в см). Три класса по 50 выборок. Setosa хорошо отделима от двух других; Versicolor и Virginica частично перекрываются.

Цель MLP: Многоклассовая классификация по ботаническим признакам. Поскольку у калькулятора лишь один выходной нейрон и нет Softmax-выхода, для каждого класса обучается отдельная сеть («класс против остальных», целевой столбец 0/1); прогнозируется класс с наибольшим выходом. Классический эталонный набор данных для сравнения алгоритмов классификации.

Параметр	Рекомендация
Архитектура	4 → 8–16 → 4–8 → 1 (на каждый класс)
Активация скрытых слоёв	ReLU
Активация выходного слоя	Sigmoid (Softmax недоступен)
Функция потерь	BCE
Скорость обучения	0,01
Эпохи	2000 – 4000
Размер батча	16 – 32
Нормализация	активна

Регрессия временных рядов (напр. температура, акции)

Несколько сдвинутых по времени значений как вход → следующее значение как выход.

Описание данных: 366 суточных температур с сезонным ходом (синусоидальный годовой цикл ± шум). Входы – температуры последних 3 дней, выход – температура следующего дня. Значения в диапазоне примерно -3 до $+23$ °C.

Цель MLP: Прогнозирование следующего значения временного ряда по прошлым наблюдениям – так называемая *авторегрессия*. Показывает, как MLP без явной памяти может использовать краткосрочные зависимости во временных рядах. Применимо к котировкам акций, энергопотреблению или данным датчиков.

Параметр	Рекомендация
Архитектура	n → 16–32 → 16 → 1
Активация скрытых слоёв	Tanh или ReLU
Активация выходного слоя	Linear
Функция потерь	MSE
Скорость обучения	0,001 – 0,005
Эпохи	3000 – 8000
Размер батча	16 – 64
Нормализация	**всегда активна**

Прогнозирование цен на жильё (несколько признаков)

Площадь, комнаты, расположение, год постройки и др. → цена. 500 выборок, 4 входа.

Описание данных: 500 синтетических записей о недвижимости с признаками: жилая площадь (40–200 м²), количество комнат (1–6), год постройки (1950–2023) и расстояние до центра города (1–30 км). Цена рассчитана как взвешенная линейная комбинация этих факторов с шумом (реалистично: 50 000–650 000 EUR).

Цель MLP: оценка непрерывной стоимости недвижимости по нескольким разнородным признакам — типичная *задача регрессии* с практическим значением. Демонстрирует работу с разномасштабными входами (поэтому нормализация обязательна) и важность разделения Train/Test для оценки обобщающей способности.

Параметр	Рекомендация
Архитектура	n → 32–64 → 16–32 → 1
Активация скрытых слоёв	ReLU
Активация выходного слоя	Linear
Функция потерь	MSE
Скорость обучения	0,001
Эпохи	3000 – 10000
Размер батча	32 – 64
Нормализация	**всегда активна**
Разбивка Train/Test	80 / 20

Обнаружение спама / Бинарная классификация текста

Векторы признаков текста (напр. частоты слов) → Спам (1) / Не спам (0). 400 выборок, 5 входов.

Описание данных: 400 синтетических представлений электронных писем с 5 нормированными признаками: частота слов *бесплатно*, *выигрыш*, *нажать*,

количество ссылок и частота восклицательных знаков. Спам-письма (200 выборок) показывают значительно более высокие значения всех признаков по сравнению с обычными письмами (200 выборок). Сбалансированное распределение классов (50/50).

Цель MLP: Бинарная классификация для автоматического обнаружения спама – *задача классификации* с прямой практической пользой. Показывает, как MLP может выработать надёжное правило решения на основе простых текстовых признаков. Применимо к другим задачам бинарной классификации: обнаружению мошенничества или медицинской диагностике.

Параметр	Рекомендация
Архитектура	n → 32–64 → 16 → 1
Активация скрытых слоёв	ReLU
Активация выходного слоя	Sigmoid
Функция потерь	BCE
Скорость обучения	0,001 – 0,01
Эпохи	2000 – 5000
Размер батча	32 – 64
Нормализация	активна (нормировать значения TF-IDF)
Разбивка Train/Test	80 / 20

Общие практические правила

Ситуация	Рекомендация
Loss расходится (NaN)	Уменьшить скорость обучения вдвое, проверить нормализацию
Loss рано перестаёт снижаться	Немного увеличить скорость обучения, больше нейронов
Loss падает, но Test-Loss растёт	Переобучение – меньше нейронов или больше данных
Выход всегда близок к 0 или 1	Неверная функция активации выхода (Linear вместо Sigmoid?)
Очень медленная сходимость	Проверить инициализацию Ксавье (в калькуляторе уже активна)
Малый объём данных (<100)	Малый размер батча (4–8), больше эпох
Сеть не лучше правила «как вчера»	Временной ряд: включить значение предыдущего дня как вход; искать входы с реальной связью с целью
Много тестовых строк с «экстраполяцией»	Выбрать более длинный обучающий период или проверить ошибки отмеченных строк
Свои данные: начальные значения	Скрытый 1 = 2–4 × входов, скрытый 2 ≈ вдвое меньше; выход Linear (регрессия) или Sigmoid (цель только 0/1)

Инициализация Ксавье и начальное значение (калькулятор v20): начальные веса выбираются равномерно в $\pm\sqrt{1/n_{\text{inputs}}}$. Это предотвращает взрыв или затухание градиентов в начале обучения. Случайные числа берутся из генератора с фиксированным начальным значением (seed, вкладка «Конфигурация»); поэтому начальные веса, порядок перемешивания и весь результат воспроизводимы.