

Нейронные сети и их практическое применение.

Лекция 6.

Практические рекомендации по использованию алгоритма обратного распространения ошибки.

Дмитрий Буряк

к.ф.-м.н

dyb04@yandex.ru

Формирование обучающей выборки

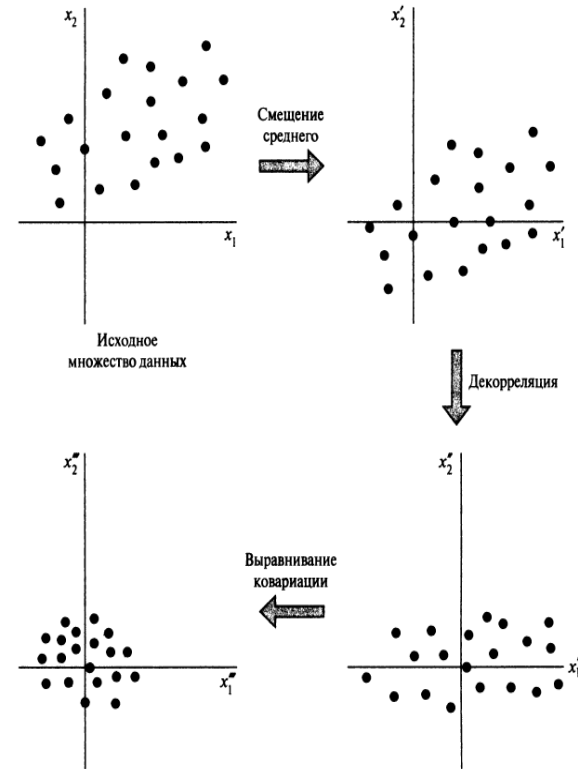
- ❑ Размер обучающей выборки
 - $N=O(|W|)$, W – множество весов. «Чем больше, тем лучше»
- ❑ Состав
 - Репрезентативность
 - Согласованность с валидационной и тестовой выборками
 - Аугментация данных:
 - добавление шумов
 - искажение данных
 - модели для генерации синтетических данных

Формирование пакетов при обучении

- Пакет – часть обучающей выборки, по которой вычисляется коррекция весов на текущем шаге.
- Принцип максимизации информативности
 - использование примеров, вызывающих наибольшие ошибки обучения
 - использование кардинально отличающихся примеров
- Случайный порядок следования примеров
- Схема акцентирования
- Коррекция распределения данных в обучающей выборке

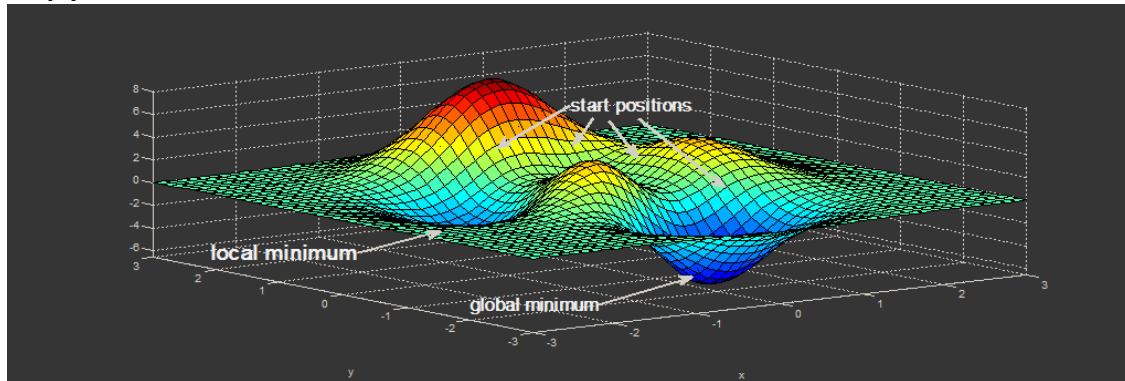
Предобработка входных данных

- Целевые значения должны находится в области значений функции активации
- Нормировка входов
 - нулевое среднее
 - отсутствие корреляции
 - одинаковая ковариация



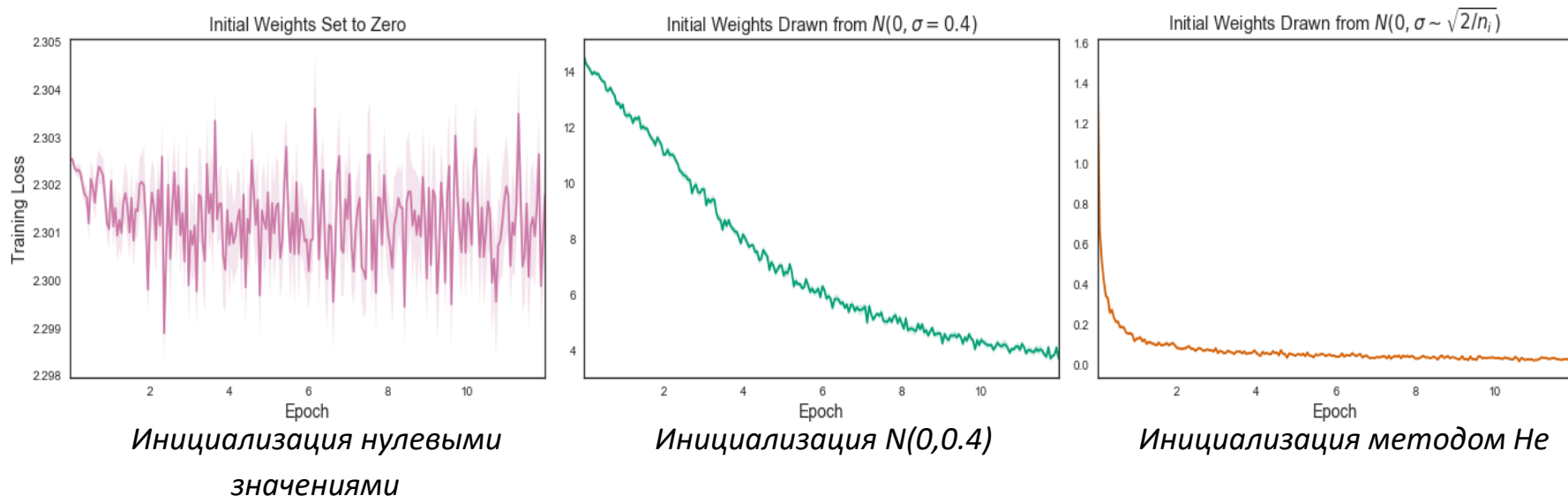
Инициализация весов

- ❑ Инициализация весов – определение начальной точки для процесса оптимизации функции ошибки .
- ❑ Основные методы основаны на эвристиках.
- ❑ Влияние на обобщающую способность НС плохо изучена.
- ❑ Обеспечение асимметричности выходов нейронов и «динамического» диапазона функции активации.



Влияние на эффективность обучения

Классификация изображений MNIST



Источник: <https://intoli.com/blog/neural-network-initialization/>

Методы инициализации

❑ Инициализировать нулями нельзя.

❑ Небольшими случайными числами.

❑ $w \in N(0, \frac{1}{\sqrt{n}}) \Rightarrow u = \sum_1^n w_i x_i \in N(0,1)$;

❑ $w \in N(0, \frac{1}{\sqrt{n_{in} + n_{out}}})$ (Glorot et al., Understanding the difficulty of training deep feedforward neural networks);

❑ $w \in N(0, \sqrt{\frac{2}{n}})$ для ReLU (He et al., Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification);

❑ Разреженная (sparse) инициализация ;

❑ Инициализация смещений – нулевые значения.

Инициализация весов. Обоснование

- Входы: нулевое среднее, единичная дисперсия, некоррелированы:

$$E[y_i y_k] = \begin{cases} 1 & \text{для } k = i \\ 0 & \text{для } k \neq i \end{cases}$$

- Начальные синаптические веса: равномерно распределены, нулевое среднее. Определить дисперсию.
- Среднее и дисперсия линейной комбинации нейронов

$$\mu_v = E[v_j] = E\left[\sum_{i=1}^m w_{ji} y_i\right] = \sum_{i=1}^m E[w_{ji}] E[y_i] = 0,$$

$$\begin{aligned} \sigma_v^2 &= E[(v_j - \mu_v)^2] = E[v_j^2] = E\left[\sum_{i=1}^m \sum_{k=1}^m w_{ji} w_{jk} y_i y_k\right] = \\ &= \sum_{i=1}^m \sum_{k=1}^m E[w_{ji} w_{jk}] E[y_i y_k] = \sum_{i=1}^m E[w_{ji}^2] = m \sigma_w^2, \end{aligned}$$

- Если $\sigma_w = m^{-1/2}$ то $\sigma_v = 1$

Перекрестная проверка

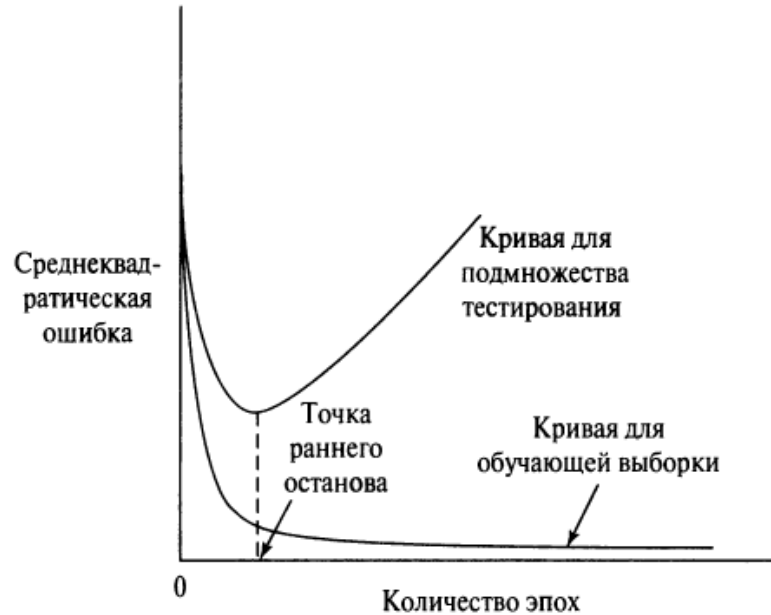
- ❑ Обучающее множество
 - обучающее подмножество
 - проверочное подмножество

- ❑ Тестовое множество

N – размер обучающего множества
 $(1-r)*N$ – размер обучающего подмножества
 $r*N$ – размер проверочного подмножества
Например: $r=0.2$

- ❑ Применение
 - Выбор оптимальной архитектуры
 - Обучение с ранним остановом
 - Многократная перекрестная проверка

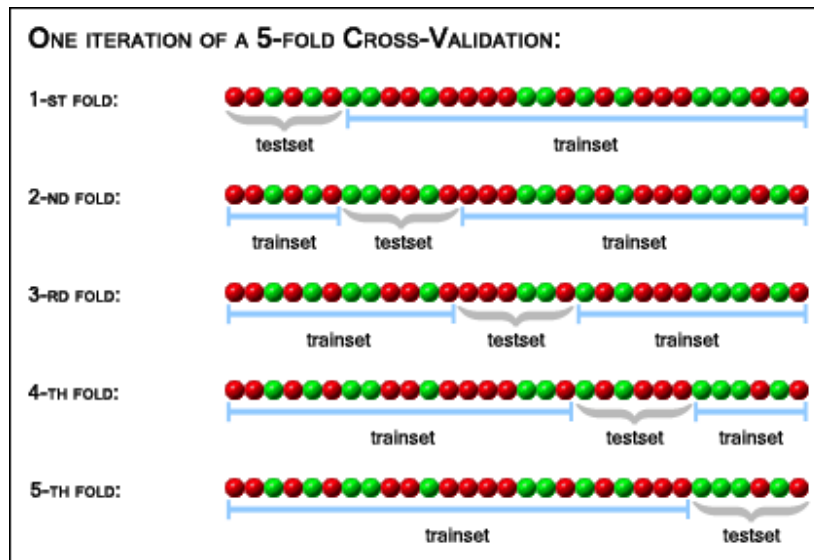
Обучение с ранним остановом



□ При $N \gg W$ эффективность падает.

Многократная перекрестная проверка

- Делим N примеров на K подмножеств
- Обучаем на $K-1$ подмножестве, тестируем на оставшемся
- Повторяем K раз
- Вычисляем среднюю ошибку по всем циклам
- Если N мало, то $K=N$.



Основные причины низкой эффективности НС

- Низкая эффективность = большая ошибка на тестовых данных.
- Проблемы с данными;
- Несоответствие архитектуры НС сложности задачи;
- Неоптимальные значения гиперпараметров;
- Переобучение;
- Ошибки в реализации.

Обозначения

- ❑ S_{train} – обучающая выборка; S_{test} – тестовая выборка;
- ❑ E_{train} – ошибка на обучающей выборке, E_{test} – ошибка на тестовой выборке, E_{goal} – целевое значение ошибки.
- ❑ $E_{test} > E_{goal}$

Анализ ошибки на обучающей выборке

$E_{train} > E_{goal}$

- Увеличить размер НС;
- Улучшить алгоритм обучения
- Оптимизировать значения гиперпараметров алгоритма обучения
- Анализ качества исходных данных
 - низкое значение сигнал-шум;
 - ошибки алгоритма предобработки;
 - недостоверные референсные значения;
 - несбалансированная выборка.

Анализ ошибки на тестовой выборке

$E_{train} < E_{goal}$ и $E_{test} > E_{goal}$

Увеличить размер S_{train} ;

Уменьшить размер НС;

Оптимизировать значения гиперпараметров НС (регуляризация);

Подбор алгоритма обучения;

Несоответствие S_{train} и S_{test} .

Вопросы

1. Основные принципы формирования обучающей выборки.
2. Основные действия для предобработки данных
3. Для чего используют многократную перекрестную проверку?