

# Машинное обучение (Machine Learning)

## Обучение на одном примере (One-shot learning)

Уткин Л.В.

Санкт-Петербургский политехнический университет Петра Великого



# One-Shot Learning

# Пример



# Формальная постановка задачи

Дано:

- малое “помеченное” обучающее множество  $S$  из  $N$  примеров единичной размерности с метками  $y$

$$S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$$

- тестовый пример  $\hat{\mathbf{x}}$ , который нужно классифицировать

Цель:

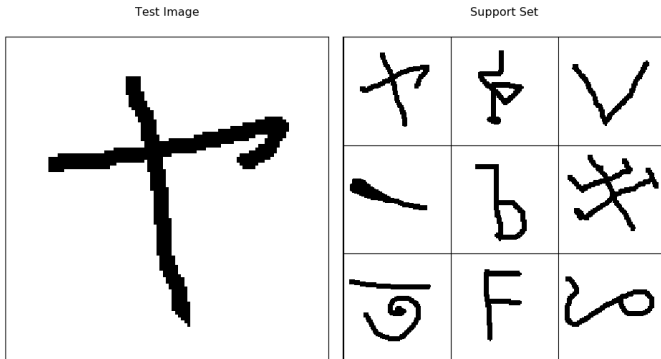
- так как ровно один пример имеет “правильный” класс, то необходимо определить  $y \in S$  такое же как метка  $\hat{y}$  примера  $\hat{\mathbf{x}}$

# Что нужно учесть при решении

- В реальности не всегда есть ограничение, что только одно изображение имеет правильный класс
- Просто обобщить эту ситуацию на случай  $k$ -shot, если есть не один, а  $k$  примеров для каждого  $y_i$ , а не один.
- Когда  $N$  большое, есть большее число возможных классов, к которым может принадлежать  $\hat{\mathbf{x}}$ , поэтому сложнее предсказать правильный класс.
- Случайное угадывание будет иметь  $\frac{100}{N}\%$  точность в среднем

# Примеры

Датасет Omniglot  $N = 9$

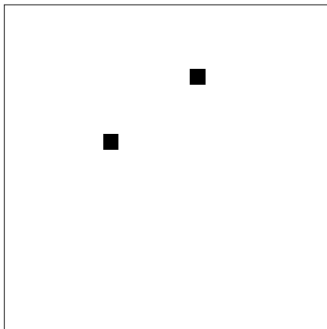


Датасет Omniglot представляет собой набор из 1623  
рисованных символов в разрешении 105x105 из 50 алфавитов.

# Примеры

Датасет Omniglot  $N = 25$

Test Image



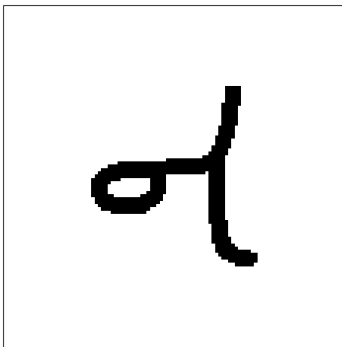
Support Set



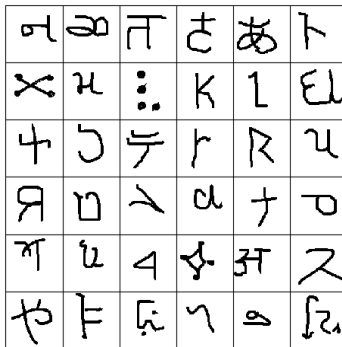
# Примеры

Датасет Omniglot  $N = 36$

Test Image



Support Set





# Omniglot

## Sanskrit

प	झ	ञ	ष	म	ल	घ
ट	ठ	क	त्र	फ	अ	व
ड	ण	न	ज	ग	थ	स
द	ओ	भ	औ	य	उ	त
र	छ	ण	इ	ल	थ	ढ
क्व	च	इ	ब	ह	श	क्व

Greek

$\varphi$	$\mathcal{L}$	$\beta$	$\delta$	$\lambda$
$\mu$	$\alpha$	$\kappa$	$\chi$	$\nu$
$\upsilon$	$\theta$	$\gamma$	$\tau$	$\sigma$
$\omega$	$\pi$	$\eta$	$\omicron$	$\epsilon$
$\rho$	$\xi$	$\zeta$	$\psi$	

Bengali

ঐ	ঈ	আ	ব	ভ	ষ	জ
ঔ	ক	খ	গ	ঘ	ঙ	চ
দ	ধ	প	ফ	ব্ৰ	ই	জ
শ	ছ	ত্ৰ	ড	ম	ন	য
র	ত	হ	ঋ	ঌ	ঔ	শ
ঢ	গ	ঢ	ল	ল্ৰ	ট্ৰ	ষ
ঠ	ফ	ধ	ব			

# Простейший метод классификации - 1 ближайший сосед

- Простейший способ классификации - это  $k$  ближайших соседей, но поскольку для каждого класса есть только один пример, используем 1 ближайшего соседа.
- Евклидово расстояние от тестового примера до обучающего:

$$C(\hat{\mathbf{x}}) = \arg \min_{c \in S} \|\hat{\mathbf{x}} - \mathbf{x}_c\|$$

- Точность (Koch и др.):  $\sim 28\%$  при  $N = 20$  omniglot
- Это примерно в 6 раз больше, чем просто случайное угадывание (5%)
- У людей точность 95.5% при  $N = 20$  omniglot
- *Hierarchical Bayesian Program Learning* (Lake и др.) дает 95.2%

# Нейронные сети для обучения

- Как обучить нейронную сеть на единичных примерах?  
Переобучение!
- Многие подходы используют Transfer Learning
- Вспомним 1 ближайшего соседа - просто классифицирует путем поиска ближайшего примера на расстоянии  $L_2$  (Евклидово расстояние)
- Но эта метрика плоха для большой размерности

# Сиамские сети



# Сиамские сети

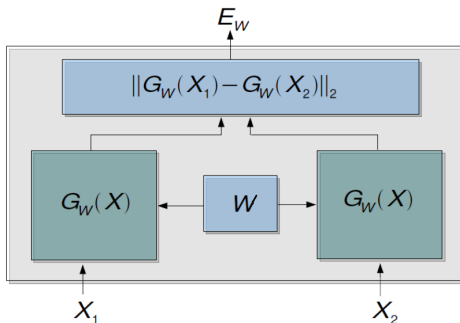
Идея: Сиамская сеть может сравнивать тестовое изображение с каждым изображением в наборе и выбирать, какое из них, имеет один и тот же класс - наиболее близко.

# Элементы сиамских сетей

- $X_1$  и  $X_2$  - пара изображений
- $Y = 0$ , если  $X_1$  и  $X_2$  - один объект,  $Y = 1$ , если  $X_1$  и  $X_2$  - различны
- Построить нейронную сеть с минимальным числом параметров, определяющую для пар объектов, одинаковы ли она или нет

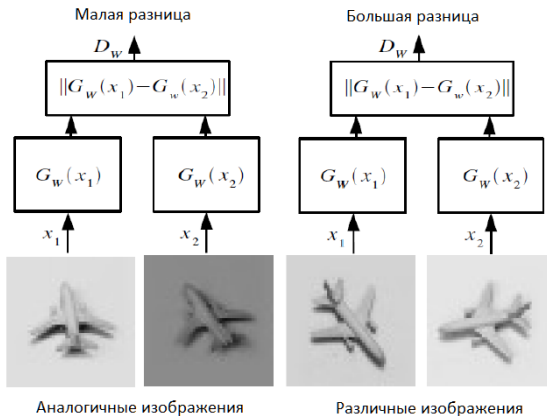
# Архитектура сиамских сетей

Y. LeCun. Learning Hierarchies of Invariant Features



- $W$  - общий вектор параметров,
- $G_W(X_1)$ ,  $G_W(X_2)$  - точки в прост-ве меньшей размерности
- $E_W$  - функция совместимости между  $X_1$  и  $X_2$  (“энергия”)

# Еще пример сиамских сетей



Y. LeCun. Learning Hierarchies of Invariant Features



# Функция потерь

- Функция потерь зависит от входных данных и параметров косвенно через энергию:

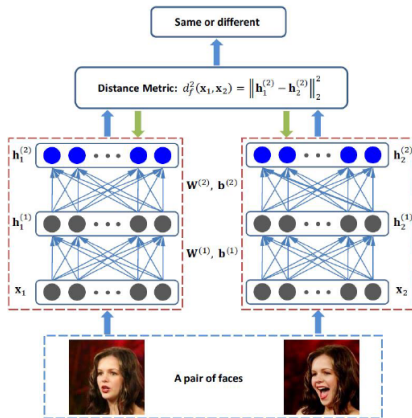
$$\mathcal{L}(W) = \sum_{i=1}^N L(W, (Y, X_1, X_2)_i)$$

$$L(W, Y, X_1, X_2) = (1-Y)L_G(E_W(X_1, X_2)) + YL_I(E_W(X_1, X_2))$$

$$E_W = \|G_W(X_1) - G_W(X_2)\|$$

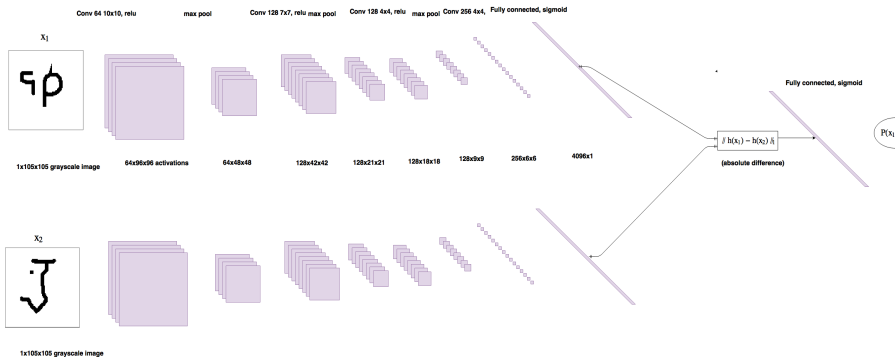
- $L_G$  - функция потерь для совпадающих пар  $Y = 0$
- $L_I$  - функция потерь для несовпадающих пар  $Y = 1$

# Применение к распознаванию лиц



JunlinHu, etc. Discriminative Deep Metric Learning for Face Verification in the Wild,  
CVPR 2014

# Глубокая сиамская сеть



# Глубокая сиамская сеть

- Используем  $t = 1$ , если два изображения одного класса и  $t = 0$  иначе
- Функция потерь

$$\begin{aligned} L(\mathbf{x}_1, \mathbf{x}_2, t) = & t \cdot \log(p(\mathbf{x}_1 \circ \mathbf{x}_2)) \\ & + (1 - t) \cdot \log(1 - p(\mathbf{x}_1 \circ \mathbf{x}_2)) \\ & + \lambda \cdot \|w\|_2 \end{aligned}$$

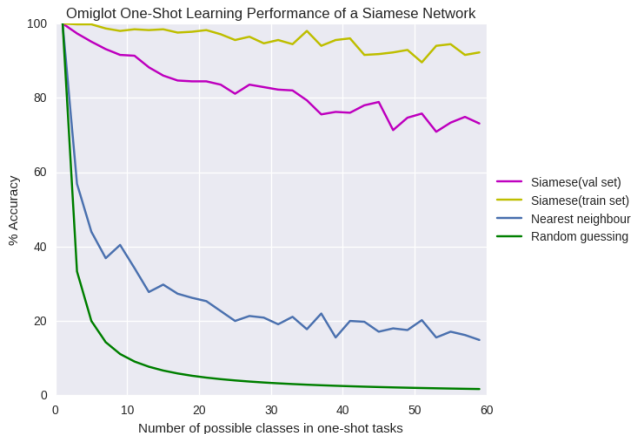
- Решение

$$C(\hat{\mathbf{x}}, S) = \arg \max_c P(\hat{\mathbf{x}} \circ x_c), \quad x_c \in S$$

# Глубокая сиамская сеть - обучение

- Почему нет переобучения
- Если есть  $C$  примеров в  $E$  классах, то число пар среди  $C \cdot E$  примеров  $N_{\text{пар}} = C \cdot E \cdot (1 - C \cdot E)/2$
- 20 примеров Omniglot из 964 классов - 185 849 560 пар!
- Но число примеров одного класса  $N_{\text{одинак}} = \binom{E}{2} C$ .  
Это 183 160 пар.
- Важно: для обучения сиамской сети необходимо соотношение 1 : 1 примеров одного и разных классов

# Характеристики



<https://sorenbouma.github.io/blog/oneshot/>

# Вопросы

?