

Сравнение методов классификации основанных на тензорных разложениях

Команда:

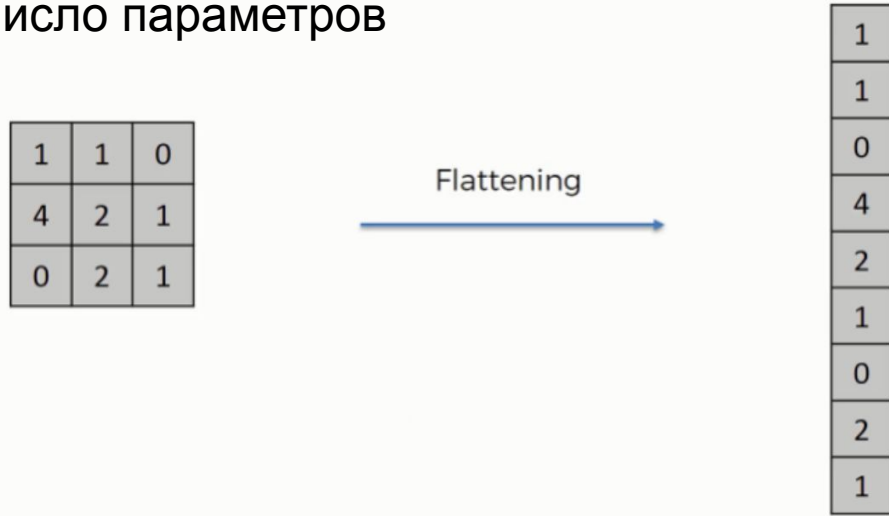
Алгебраический эскадрон

Состав:

- Сергей Грозный
- Роман Кравченко
- Дмитрий Шмыголь

Введение

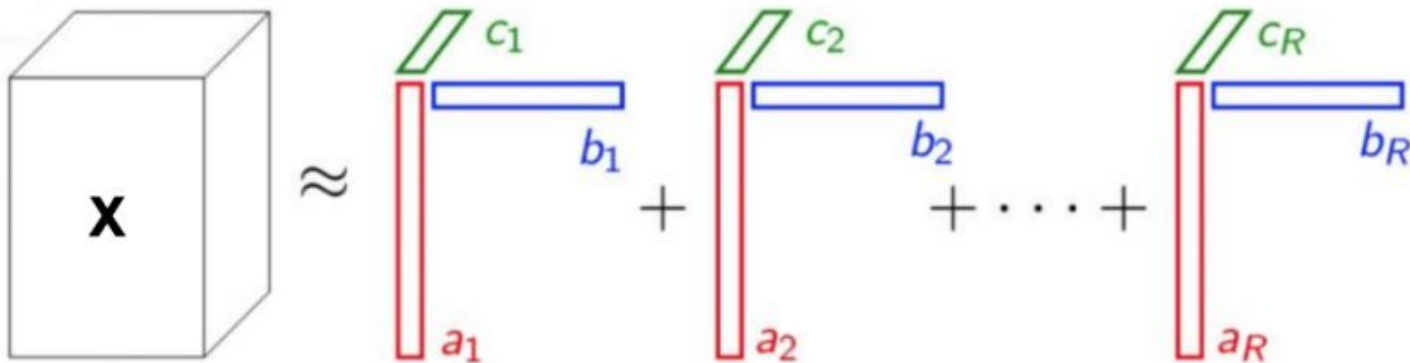
- Обучающие данные часто имеют тензорную структура
- Стандартный подход: вытягивание данных в вектор
- Минусы:
 - теряем пространственную и временную информацию
 - большое число параметров



Введение

- Хотим сразу обучаться на входном тензоре используя CP разложение:
- $\mathbf{X} \in \mathbb{R}^{D_1 \times D_2 \times D_3}$

$$\mathbf{X} \approx \sum_{r=1}^R a_r \circ b_r \circ c_r$$



Постановка задачи

- Изучение, реализация и сравнение линейных методов классификации основанных на CP разложении
- Измерение качества:

$$\text{MSE} = \frac{1}{n} \|W - \hat{W}\|^2$$

$$\cos(\theta) = \frac{\langle W \cdot \hat{W} \rangle}{\|W\| \cdot \|\hat{W}\|}$$

$$\text{Accuracy} = \frac{\# \text{ of correct predictions}}{\text{total } \# \text{ of predictions made } (n)}.$$

SVM

- Стандартный SVM:

$$\underset{w}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \max[0, 1 - y_i(w^\top x_i)] + \frac{\lambda}{2} \|w\|^2$$

- Рассматриваемый CP-SVM:

$$\underset{W_1, \dots, W_N}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \max[0, 1 - y_i(\langle \sum_{r=1}^R W_1^{(r)} \circ W_2^{(r)} \circ \dots \circ W_N^{(r)}, \mathbf{X}_i \rangle)]$$

Logistic Regression

- Стандартный Logistic Regression:

$$\underset{w}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(w^\top x_i))) + \lambda \|w\|^2.$$

- Рассматриваемый CP-LogReg:

$$\underset{W_1, \dots, W_N}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(\langle \sum_{r=1}^R W_1^{(r)} \circ W_2^{(r)} \circ \dots \circ W_N^{(r)}, \mathbf{X}_i \rangle)))$$

Поиск минимума

- Используется следующий алгоритм:

Algorithm 1 CP Alternating Minimization (Zhou et al.)

Require: Dataset $\{(\mathbf{X}_i, y_i)\}_{i=1}^n$ with $\mathbf{X} \in \mathbb{R}^{D_1 \times \dots \times D_N}$, $y_i \in \{-1, 1\}$ and R

1: Initialize: $A_i \in \mathbb{R}^{D_i \times R}$ for $i = 1, \dots, N$

2: **repeat**

3: **for** $i = 1, \dots, N$ **do**

4: $A_i^{(t+1)} = \underset{A_i}{\operatorname{argmin}} \ell(\mathbf{X}, y, A_1^{(t+1)}, \dots, A_{i-1}^{(t+1)}, A_i, A_{i+1}^{(t)}, \dots, A_N^{(t)}) + \lambda \|A_i\|^2$

5: **end for**

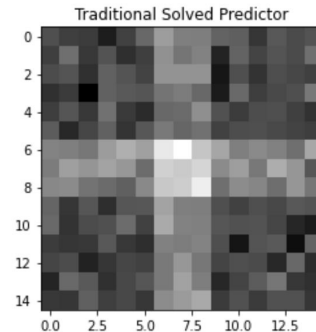
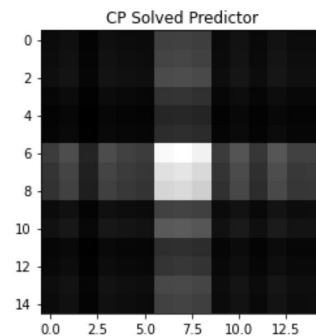
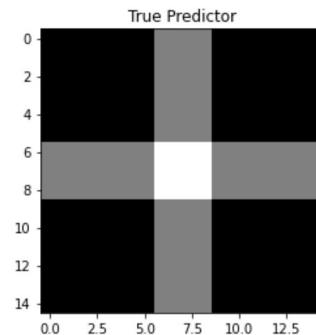
6: **until** $\ell(\theta^{(t+1)}) - \ell(\theta^{(t)}) < \epsilon$

Эксперименты

- Сгенерированные данные:

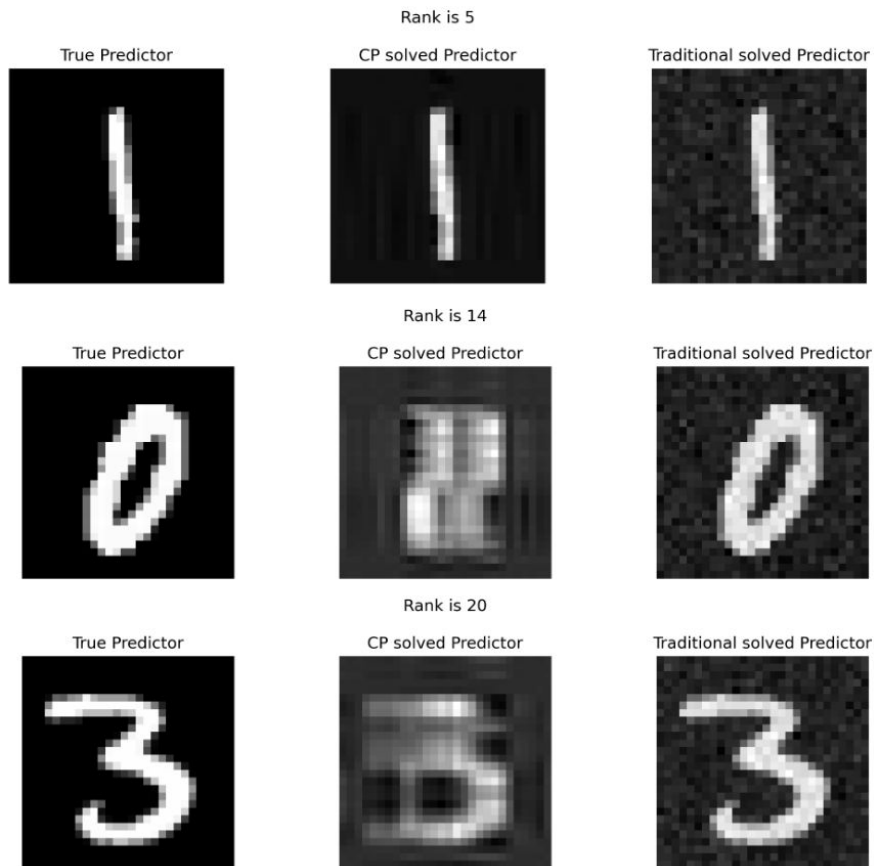
- $y_i = \langle X_i, W \rangle + \epsilon_i$
- ϵ_i, X_i из $\mathcal{N}(0,1)$
- W - True Predictor

Метод	MSE	Число параметров
CP Solved Predictor	0.15	60
Traditional Solved Predictor	0.18	225



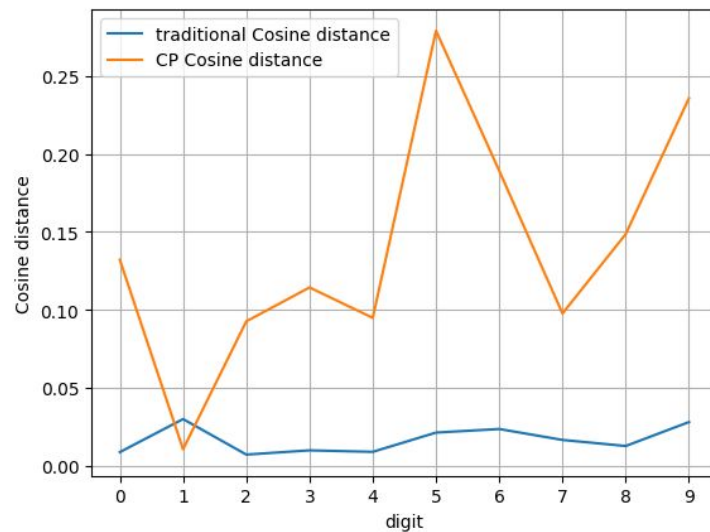
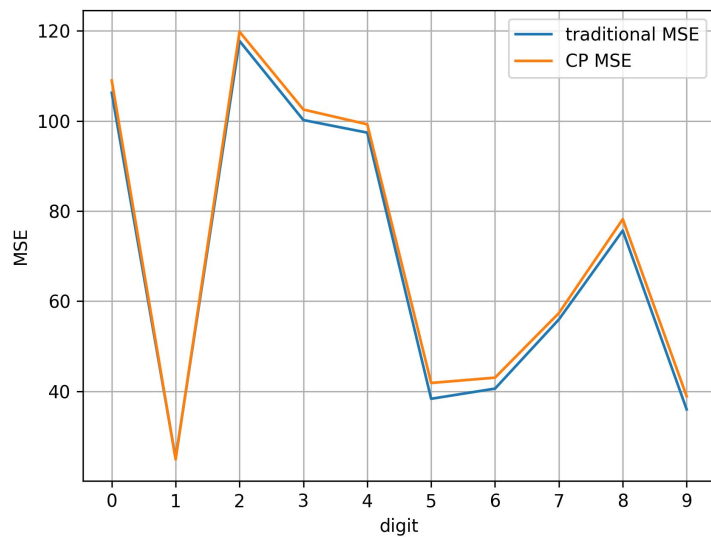
Эксперименты

- Сгенерированные данные используя MNIST:
 - $y_i = \langle X_i, W \rangle + \epsilon_i$
 - ϵ_i, X_i из $\mathcal{N}(0,1)$
 - W - картинка из MNIST



Эксперименты

- Сравнения для всех цифр:



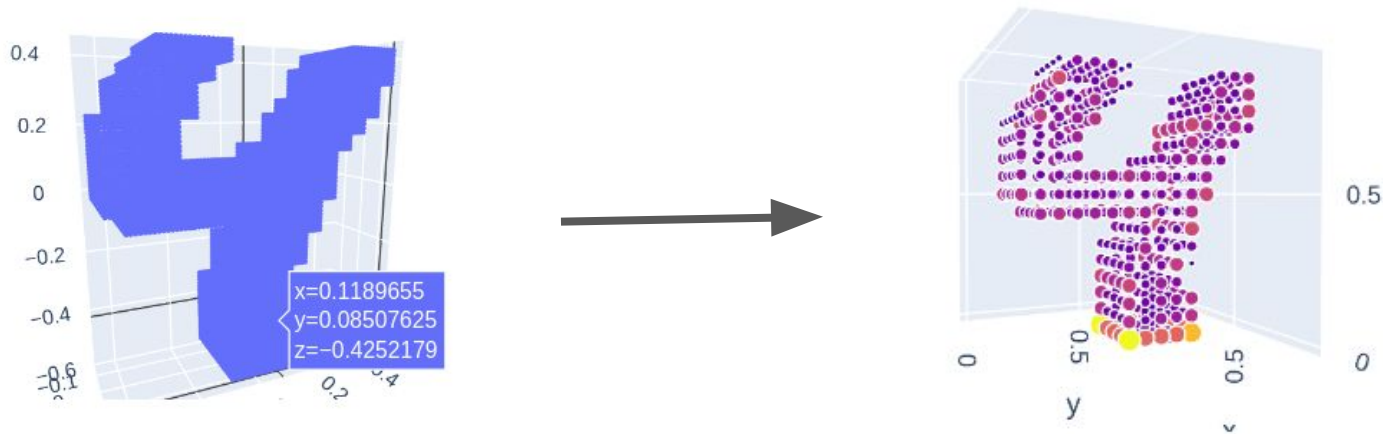
Результаты

- MNIST:
 - обучение: 12600 изображений
 - тестирование: 3100 изображений

Метод	sample shape	Точность	Время, с	Параметры
Log Regression	784×1	0.98	7	784
CP Log Regression	28×28	0.93	25	112
SVM	784×1	0.98	105	784
CP SVM	28×28	0.96	186	112

Эксперименты

- Используется 3D-MNIST, понизив размерность с ~20000 точек до размера $16 \times 16 \times 16$:



Результаты

- 3D MNIST:
 - обучение: 100 изображений
 - тестирование: 100 изображений

Метод	sample shape	Точность	Время, с	Параметры
Log Regression(sklearn)	4096×1	0.86	—	4096
CP Log Regression	16×16×16	0.87	102	144
SVM	4096×1	0.86	196	4096
CP SVM	16×16×16	0.74	9	144

Выводы

- Тензорные методы позволяют использовать в разы меньше параметров
- На MNIST тензорные методы проигрывают обычным алгоритмам, но на 3D-MNIST уже выигрывают
- Стоит смотреть в сторону тензорных методов, если:
 - искомые веса имеют низкоранговую структуру
 - обучающие данные - это многомерные тензоры

ОСНОВНЫЕ ССЫЛКИ

- “Learning Predictors from Multidimensional Data with Tensor Factorizations”
Soo Min Kwon, Anand D. Sarwate
- “Tensor-Train decomposition for image classification problems”
- “Low-Rank Tensor Networks for Dimensionality Reduction and Large-Scale Optimization Problems: Perspectives and Challenges”
- “Тензоризованные нейронные сети” - Гарипов Тимур Исмагилевич

Github:

https://github.com/PotencialGibsa/NLA_project2.git