

Регуляризация нейросетевого слоя путем построения фрейма в пространстве параметров

Григорьев Алексей Дмитриевич

Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Кафедра интеллектуальных систем

Научный руководитель: к.ф.-м.н. А.Н. Гнеушев

Москва, 2022

Задача

Предложить метод увеличения эффективности избыточного множества параметров нейронной сети и повышения устойчивости модели.

Проблема

Существующие решения, предполагающие регуляризацию параметров модели, накладывают чрезмерные ограничения на оптимизацию весов нейросети, что негативно влияет на качество модели.

Решение

Рассматривать веса слоя нейросети как систему векторов, проекция входа на которую устойчива и полна. Допустимая избыточность данной системы свойственна нейронной сети и позволяет точнее описывать ее веса.

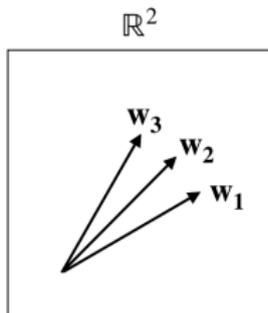
Прунинг параметров

- *P. Molchanov, et al.* Pruning convolutional neural networks for resource efficient transfer learning // ICLR, 2017, P. 1–17.

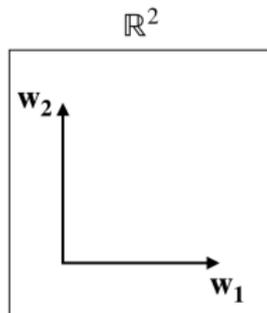
Повышение разнообразия нейронов

- *W. Lui, et al.* Learning towards minimum hyperspherical energy // NIPS, 2018, P. 6222–6233.
- *N. Bansal, et al.* Can we gain more from orthogonality regularizations in training deep CNNs? // NIPS, 2018, P. 4266–4276.
- *J. Wang, et al.* Orthogonal Convolutional Neural Networks // CVPR, 2020, P. 11505–11515.

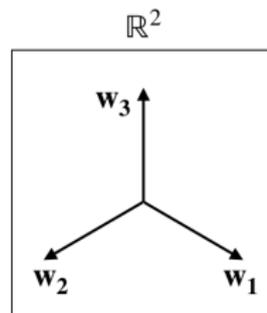
- Коррелированные системы весов нейронов неэффективны.
- Ортогональность – чрезмерное требование и ограничение.
- Избыточные полные системы могут быть адекватны.
- Предлагается построение полной системы для разложения входных векторов в избыточном пространстве весов каждого слоя.



*коррелированная
избыточная
система*



*ортогональная
система*



*адекватная
избыточная
система*

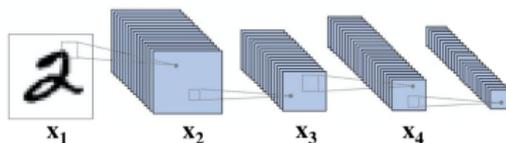
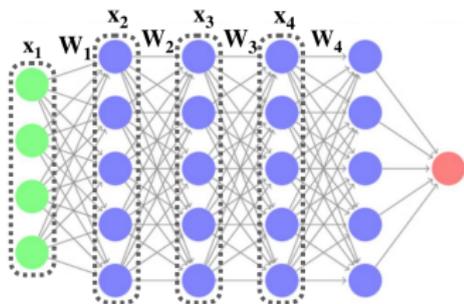
Возможные конфигурации весов нейронов

Семейство моделей

- Ограничим множество рассматриваемых моделей семейством Φ_L нейросетей следующего вида.
- $\varphi(\cdot|\Theta) \in \Phi_L$ – модель из семейства глубоких нейронных сетей, состоящих из L слоев, каждый из которых представим в виде линейного оператора $\mathcal{F}_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{m_i}$ и нелинейной функции активации $h : \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{m_i}$:

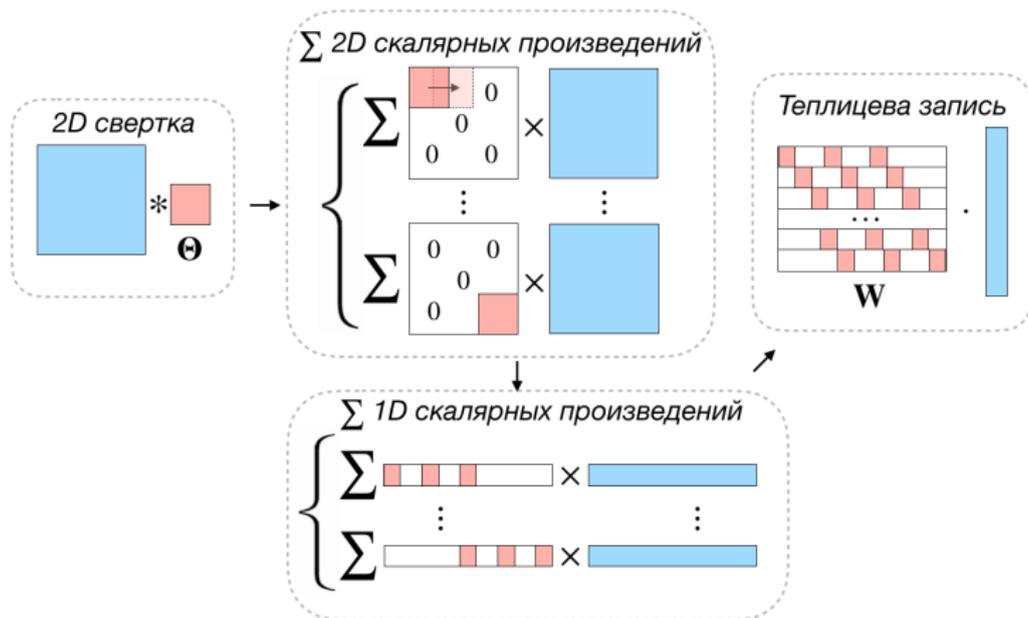
$$h(\mathcal{F}_i(\mathbf{z})) = h(\mathbf{W}_i \mathbf{z}), \quad \forall \mathbf{z} \in \mathbb{R}^{n_i}, \quad i = 1, \dots, L,$$

где $\mathbf{W}_i \in \mathbb{R}^{m_i \times n_i}$, $\mathbf{W}_i \subseteq \Theta$ – матрица линейного оператора, составленная из параметров данного слоя, $n_i, m_i : m_i \geq n_i$ – размерности входа и выхода слоя соответственно.



Линейное представление сверточного слоя

Сверточный слой в виде линейного оператора с матрицей \mathbf{W} задается блочно-теплицевой матрицей из параметров Θ свертки.



Теплицево представление одноканальной свертки

Оптимизация параметров модели

- $\{\mathbf{x}_i, y_i\}_{i=1}^N$ – выборка размера N ;
- $\varphi(\cdot|\Theta) \in \Phi_L$ – модель из семейства Φ_L глубоких нейронных сетей, состоящих из L слоев, представимых в виде линейного оператора;
- минимизация эмпирического риска:

$$\hat{\Theta} = \arg \min_{\Theta} \frac{1}{N} \sum_{i=1}^N \ell(\varphi(\mathbf{x}_i|\Theta), y_i) + \gamma \tilde{R}(\Theta),$$

где ℓ – функция потерь, релевантная задаче обучения с учителем;
 $\tilde{R}(\Theta) = \sum_{i=1}^L R(\mathbf{W}_i)$ – регуляризация, γ – коэф. регуляризации;

Регуляризация параметров

Регуляризация параметров $\mathbf{W}^T = [\mathbf{w}_1 \dots \mathbf{w}_m]$ направлена на минимизацию потерь информации на слое $\mathcal{F}(\mathbf{z}) = \mathbf{W}\mathbf{z}$ путем построения системы весов $\{\mathbf{w}_i\}_{i=1}^m$, линейно восстанавливающих вход \mathbf{z} по выходу $\mathcal{F}(\mathbf{z})$:

$$\forall \mathbf{z} \in \mathbb{R}^n \exists \tilde{\mathbf{c}} = \tilde{\mathbf{c}}(\mathcal{F}(\mathbf{z}), \mathbf{W}) : \mathbf{z} \approx \hat{\mathbf{z}} = \sum_{k=1}^m \tilde{c}_k \mathbf{w}_k.$$

Определение (фрейм)

$\{\mathbf{w}_k\}_{k=1}^m \subset \mathbb{R}^n$ – фрейм в \mathbb{R}^n , если $\exists A, B : 0 < A \leq B < \infty : \forall \mathbf{z} \in \mathbb{R}^n$ выполнено нер-во фрейма:

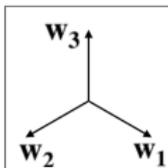
$$A\|\mathbf{z}\|^2 \leq \sum_{i=1}^m |\langle \mathbf{z}, \mathbf{w}_i \rangle|^2 \leq B\|\mathbf{z}\|^2$$

где A, B – границы фрейма. Если $A = B$, то фрейм называется жестким.

Разложение по дуальной системе

Если $\{\mathbf{w}_k\}_{k=1}^m$ – фрейм в \mathbb{R}^n , то разложение по дуальному фрейму $\{\tilde{\mathbf{w}}_i\}_{i=1}^m$:

$$\mathbf{z} = \sum_{i=1}^m \langle \mathbf{z}, \mathbf{w}_i \rangle \tilde{\mathbf{w}}_i, \quad \forall \mathbf{z} \in \mathbb{R}^n.$$



Пример жесткого фрейма с границей $A = \frac{3}{2}$ в \mathbb{R}^2

Свойства фрейма $\{\mathbf{w}_k\}_{k=1}^m \subset \mathbb{R}^n$

- Фрейм образует полную систему в \mathbb{R}^n . При $m > n$ система избыточна, что характерно для слоя нейросети и позволяет точнее его описывать.
- Если строки $\{\mathbf{w}_k\}_{k=1}^m$ матрицы \mathbf{W} образуют фрейм, то собственные числа $\lambda_1, \dots, \lambda_n$ матрицы $\mathbf{W}^T \mathbf{W}$ ограничены границами фрейма:

$$A \leq \lambda_i \leq B, \quad \forall i = 1, \dots, n.$$

- Для переопределенной СЛАУ $\mathcal{F}(\mathbf{z}) = \mathbf{W}\mathbf{z}$ фрейм $\{\mathbf{w}_k\}_{k=1}^m$ дает устойчивое решение задачи восстановления входа:
 $\mathbf{z} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathcal{F}(\mathbf{z})$. Обусловленность задачи ограничена:

$$\kappa(\mathbf{W}) = \|\mathbf{W}^T \mathbf{W}\| \|(\mathbf{W}^T \mathbf{W})^{-1}\| = \frac{|\lambda_{\max}|}{|\lambda_{\min}|} \leq \frac{B}{A}.$$

Модель слоя: $\mathbf{W} \in \mathbb{R}^{m \times n} : m \geq n$

- Нейросетевого слой задан линейным оператором $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ с матрицей $\mathbf{W} \in \mathbb{R}^{m \times n} : m \geq n$ и нелинейной функцией активации h , $h(\mathcal{F}(\mathbf{z})) = h(\mathbf{W}\mathbf{z})$, $\forall \mathbf{z} \in \mathbb{R}^n$; $\mathbf{W}^T = [\mathbf{w}_1 \dots \mathbf{w}_m]$.
- Для обратимости линейной части слоя $\mathcal{F}(\mathbf{z}) = \mathbf{W}\mathbf{z}$ необходимо и достаточно, чтобы строки $\{\mathbf{w}_k\}_{k=1}^m$ матрицы \mathbf{W} образовывали фрейм.

Построение фрейма

- Неравенство фрейма для строк $\{\mathbf{w}_k\}_{k=1}^m$ матрицы \mathbf{W} :

$$A\|\mathbf{z}\|^2 \leq \|\mathbf{W}\mathbf{z}\|^2 \leq B\|\mathbf{z}\|^2, \forall \mathbf{z} \in \mathbb{R}^n \iff \begin{cases} (\mathbf{W}^T\mathbf{W} - A\mathbb{I}) \succeq 0, \\ (-\mathbf{W}^T\mathbf{W} + B\mathbb{I}) \succeq 0. \end{cases}$$

- Матрица $\mathbf{V} \in \mathbb{R}^{m \times m}$ положительно полуопределена, если:

- она обладает свойством диагонального преобладания:

$$|v_{ii}| \geq \sum_{j \neq i} |v_{ij}| \quad \forall i = 1, \dots, m,$$

- ее диагональные элементы неотрицательны:

$$v_{ii} \geq 0 \quad \forall i = 1, \dots, m.$$

- Пусть $\mathbf{V} = \mathbf{W}^T\mathbf{W}$, $M(v) = \min(v, 0)$; введем регуляризатор:

$$R(\mathbf{W}) = \frac{1}{n} \sum_{i=1}^n \underbrace{M(v_{ii} - A - \sum_{j=1}^n |v_{ij}|)^2}_{\text{штраф } i\text{-ой строки } (\mathbf{W}^T\mathbf{W} - A\mathbb{I})} + \underbrace{M(-v_{ii} + B - \sum_{j=1}^n |v_{ij}|)^2}_{\text{штраф } i\text{-ой строки } (-\mathbf{W}^T\mathbf{W} + B\mathbb{I})}.$$

Цель

Сравнить предложенный подход к регуляризации параметров модели с существующими решениями в задаче классификации изображений.

Параметры эксперимента

- задача многоклассовой классификации;
- архитектуры модели $\varphi(\cdot|\Theta)$ – ResNet-34, ResNet-50;
- функция потерь ℓ – кросс-энтропия;
- оптимизатор – Adam с начальным шагом 0.01;
- критерий качества – Accuracy.

Выборки

- 1 CIFAR-10, CIFAR-100, SVHN – датасеты изображений;

Выборка	Число изображений	Число классов
CIFAR-10	60000	10
CIFAR-100	60000	100
SVHN	~100000	10

Ассурасу (%) методов регуляризации (ResNet-34)

Метод регуляризации	CIFAR-10	CIFAR-100	SVHN
Без регуляризации	94.53 ± 0.03	75.58 ± 0.08	96.50 ± 0.03
Minimum Hyperspherical Energy	94.58 ± 0.04	75.78 ± 0.08	96.59 ± 0.03
Weights Orthogonalization	94.59 ± 0.04	75.98 ± 0.08	96.51 ± 0.02
Spectral Restricted Isometry	94.72 ± 0.03	76.24 ± 0.09	96.57 ± 0.03
Orthogonal Convolutions	95.03 ± 0.04	76.57 ± 0.06	96.66 ± 0.02
Фреймовая регуляризация	95.17 ± 0.05	77.61 ± 0.07	96.85 ± 0.02

Ассурасу (%) методов регуляризации (ResNet-50)

Метод регуляризации	CIFAR-10	CIFAR-100	SVHN
Без регуляризации	94.83 ± 0.04	77.20 ± 0.07	96.92 ± 0.03
Minimum Hyperspherical Energy	94.88 ± 0.03	77.34 ± 0.06	96.94 ± 0.02
Weights Orthogonalization	94.92 ± 0.04	77.38 ± 0.06	96.91 ± 0.03
Spectral Restricted Isometry	95.01 ± 0.03	77.40 ± 0.07	96.95 ± 0.03
Orthogonal Convolutions	95.29 ± 0.03	77.77 ± 0.07	97.01 ± 0.02
Фреймовая регуляризация	95.25 ± 0.04	78.35 ± 0.06	97.10 ± 0.02

- Обучающая выборка CIFAR-10;
- Тестовые домены:
 - 1 CIFAR-10-C – аугментированная выборка CIFAR-10,
 - 2 CINIC-10 – подвыборка ImageNet, включающая классы из CIFAR-10;
- Для моделей с регуляризацией выбраны субоптимальные эпохи;

Ассурасу (%) методов регуляризации на разных доменах (ResNet-34)

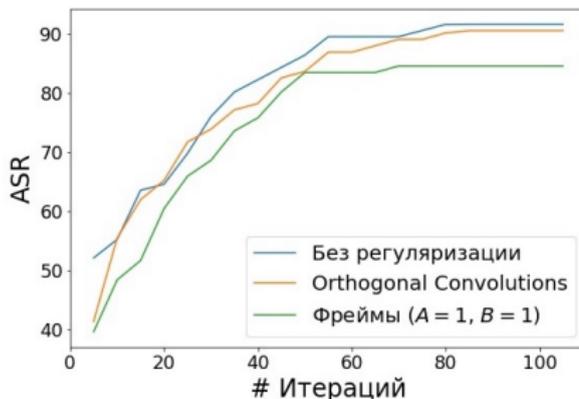
Метод регуляризации	CIFAR-10 (*)	CIFAR-10-C	CINIC-100
Без регуляризации	94.53 ± 0.03	74.77 ± 0.25	67.91 ± 0.35
Orthogonal Convolutions	94.52 ± 0.01	76.27 ± 0.19	69.87 ± 0.29
Фреймовая регуляризация	94.53 ± 0.01	76.65 ± 0.15	71.20 ± 0.32

(*) – исходный домен

- Выборка CIFAR-10;
- Состязательная атака типа "черный ящик" SimBA (Guo, 2019);
- Attack Success Rate (ASR) – доля успешных атак;

Зависимость ASR (%) от числа итераций SimBA

Метод регуляризации	# Итераций				
	1	10	50	100	1000
Без регуляризации	52.08	59.37	84.38	92.71	93.75
Orthogonal Convolutions	41.30	57.61	83.69	91.30	92.06
Фреймовая регуляризация	39.56	49.45	80.20	84.61	86.81



- Предложена модель нейросетевого слоя на основе фрейма в пространстве параметров, минимизирующая потерю информации на линейном слое.
- Предложенная модель обобщена на сверточные слои с использованием блочно-теплицева представления свертки.
- Построен фреймовый регуляризатор параметров нейросетевого слоя путем введения штрафа за нарушение фреймового неравенства.
- Проведенные вычислительные эксперименты показали значимое преимущество предложенного метода и увеличение эффективности регуляризации в задачах классификации, повышения устойчивости к состязательным атакам и к смене домена по сравнению с существующими подходами.
- Предложенная регуляризация позволила отказаться от стандартной регуляризации weight decay путем введения штрафа на соблюдение верхней границы фрейма.

Публикации по теме диссертации

- Григорьев А.Д., Гнеушев А.Н. Регуляризация параметров нейронной сети на основе неравенства Рисса // Математические методы распознавания образов: Тезисы докладов 20-й Всероссийской конференции с международным участием, г. Москва 2021 г. — М.: Российская академия наук, 2021. — С. 121-122.

Публикации

- Grigorev A., Gneushev A., Litvinchev I. Re-identification-based models for multiple object tracking // Artificial Intelligence in Industry 4.0 and 5G Technology. — Hoboken: Wiley, 2022 (in press).
- Григорьев А.Д., Гнеушев А.Н. Ре-идентификация с пред-фильтрацией по качеству изображений в задаче слежения за множеством объектов // Информационные технологии, Москва: Новые технологии, 2021. — Т. 27, N 8. — С. 409–418.
- Григорьев А.Д., Гнеушев А.Н. Слежение за множеством объектов на видео изображениях с помощью ре-идентификации с предфильтрацией дескрипторов по качеству // Интеллектуализация обработки информации: Тезисы докладов 13-й Международной конференции, г. Москва 2020 г. — М.: Российская академия наук, 2020. — С. 433-434.