Skolkovo Institute of Science and Technology

Специальность: 1.2.2
Математическое моделирование, численные методы и комплексы программ

Параметрические методы вычисления оптимальных транспортных расстояний, отображений и барицентров

Соискатель: Коротин А.А.

Научный руководитель: д.ф.-м.н., доцент Бурнаев Е.В.

2022

Overview

Характеристика работы

- Актуальность исследования
- Постановка задач исследования
- Публикации и результаты

Содержание работы

- \- Численный метод для моделирования Васерштейн-2 отображений (гл. 2)
- \- Численный метод для Васерштейн-2 барицентра (гл. 3)
- \- Численный метод для моделирования Васерштейн-2 градиентного потока (гл. 4)
- \- Методология построения эталонных пар для Васерштейн-2 (гл. 5)

Заключение

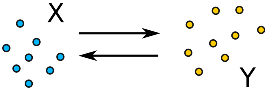
Характеристика работы

Сценарии моделирования данных в машинном обучении

Сценарий 1

Отображение

2 выборки



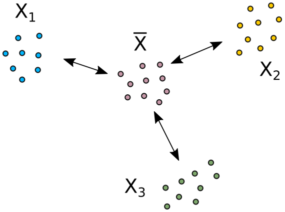
Примеры приложений:

синтез искусственных данных, доменная адаптация, сверхразрешение и перенос стиля м/у изображениями

Сценарий 2

Усреднение

Набор выборок



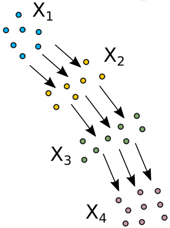
Примеры приложений:

смешивание апостериорных распределений, агрегация вероятностных прогнозов

Сценарий 3

Моделирование динамики

Послед. выборок

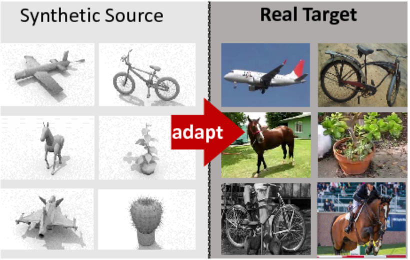


Примеры приложений:

моделирование эволюционных процесов в машинном обучении, физике, экономике

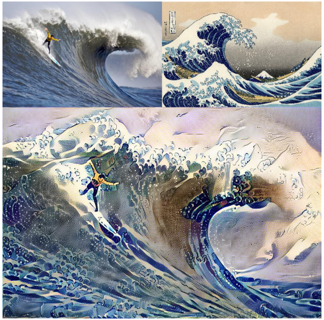
Пример сценария 1, требующего сопоставления выборок[1]

Доменная адаптация



X, Y - наборы данных исходного/целевого доменов

Передача цвета м/у изображениями

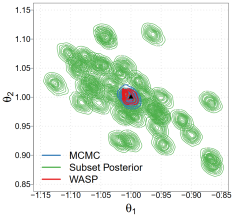


X, Y - наборы RGB пикселей

[1] Источники изображений: ai.bu.edu/syn2real/, github.com/VinceMarron/style_transfer

Пример сценария 2, требующего усреднения выборок[2,3]

Агрегация выборочных апостериорных распределений



$$p(\theta|D) \approx \text{Barycenter}\left[p(\theta|D_1), \ldots, p(\theta|D_N)\right]$$

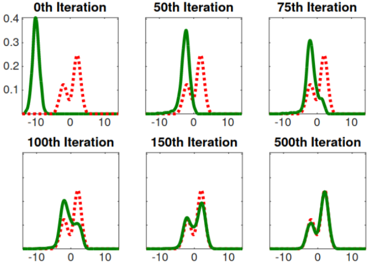
Смешивание стилей изображений

[2]Sanvesh Srivastava, Cheng Li и David B Dunson (2018). “Scalable Bayes via barycenter in Wasserstein space”. В: The Journal of Machine Learning Research 19.1, с. 312—346.
[3]Youssef Mroueh (2020). “Wasserstein Style Transfer”. В: International Conference on Artificial Intelligence and Statistics. PMLR, с. 842—852.

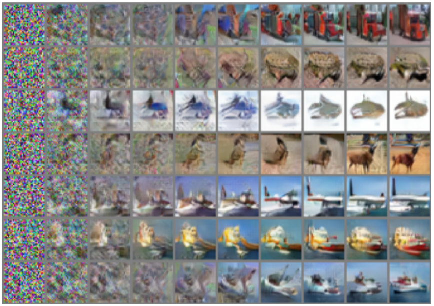
Пример сценария 3, требующего моделирования динамики[4][5]

Моделирование выборки из апостериорного распределения

$$p(\theta|D) \propto p(D|\theta)p(\theta)$$



Синтез искусственных выборок данных



[4] Qiang Liu и Dilin Wang (2016). “Stein variational gradient descent: A general purpose bayesian inference algorithm”. В: Advances in neural information processing systems 29.

[5] Jiaojiao Fan, Amirhossein Taghvaei и Yongxin Chen (2022). Variational Wasserstein gradient flow. URL: https://openreview.net/forum?id=WZR7ckBkzPY.

Сценарии моделирования данных в машинном обучении

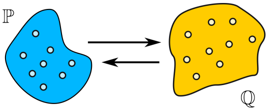
Предположение: наблюдаемые данные являются случайными выборками из соответствующих вероятностных распределений.

Сценарий 1

Отображение

2 выборки

$X \sim \mathbb{P}, Y \sim \mathbb{Q}$

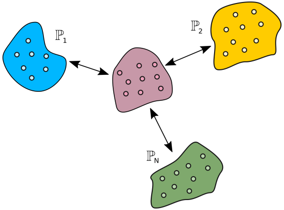


Сценарий 2

Усреднение

Набор выборок

$X_n \sim \mathbb{P}_n$

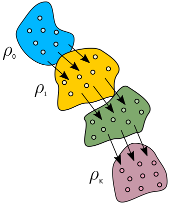


Сценарий 3

Моделирование динамики

Послед. выборок

$X_t \sim \rho_t$



Диссертационная работа[6]

<u>В данной диссертационной работе</u>
развивается арсенал математических методов моделирования для решения ранее упомянутых задач на основе подходов теории оптимального транспорта.

Задача оптимальной транспортировки массы Монже-Канторовича.

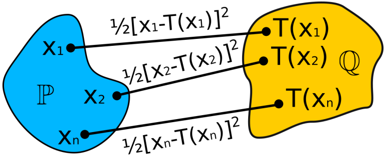
Леонид Витальевич Канторович, советский математик и экономист.

[6] Рис.: microsoft.com/en-us/research/blog/measuring-dataset-similarity-using-optimal-transport/

Задача 1: оптимальный транспорт с квадратичной ценой[7] (по Монже)

Квадрат расстояния Васерштейн-2 между $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^D)$ определяется как

$$\mathbb{W}_2^2(\mathbb{P}, \mathbb{Q}) \stackrel{\text{def}}{=} \min_{T\sharp\mathbb{P}=\mathbb{Q}} \int_{\mathbb{R}^D} \frac{\|x - T(x)\|^2}{2} d\mathbb{P}(x).$$



Функция $T^* : \mathbb{R}^D \to \mathbb{R}^D$, доставляющая минимум, называется OT отображением.
В случае $\mathbb{P} \in \mathcal{P}_{2,ac}(\mathbb{R}^D)$ T^* существует, единственно и является градиентном $\nabla\psi^*$ некоторой выпуклой функции $\psi^* : \mathbb{R}^D \to \mathbb{R}$ (Теорема Бренье).

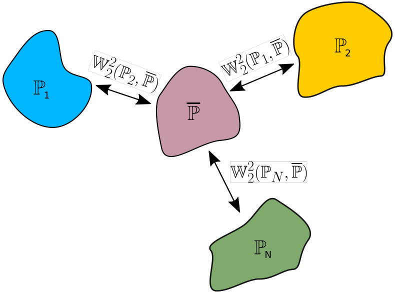
Задача 1: Вычисление оптимального отображения и расстояния Васерштейна-2.

[7] Cédric Villani (2008). Optimal transport: old and new. Т. 338. Springer Science & Business Media.

Задача 2: Барицентры в Васерштейн-2 пространстве распределений[8]

Барицентр (центр масс) $\overline{\mathbb{P}}$ распределений $\mathbb{P}_1, \mathbb{P}_2, \cdots, \mathbb{P}_N \in \mathcal{P}_2(\mathbb{R}^D)$ с весами $\alpha_1, \ldots, \alpha_N \geq 0$ при $\sum_{n=1}^{N} \alpha_n = 1$ определяется как взвешенное среднее по Фреше в пр-ве $(\mathcal{P}_2(\mathbb{R}^D), \mathbb{W}_2)$:

$$\overline{\mathbb{P}} = \operatorname*{arg\,min}_{\mathbb{P} \in \mathcal{P}_2(\mathbb{R}^D)} \sum_{n=1}^{N} \alpha_n \mathbb{W}_2^2(\mathbb{P}, \mathbb{P}_n).$$



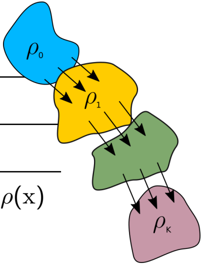
Задача 2: Вычисление Васерштейн-2 барицентра семейства распределений.

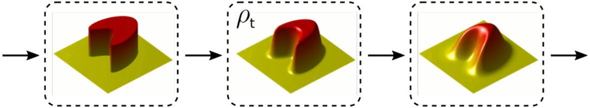
[8] Martial Agueh и Guillaume Carlier (2011). “Barycenters in the Wasserstein space”. В: SIAM Journal on Mathematical Analysis 43.2, с. 904—924.

Задача 3: Васерштейн-2 градиентные потоки[9,10]

Многие ДУЧП являются $\mathcal{W}_2 = \sqrt{2}\mathbb{W}_2$ град. потоками функционалов:

Класс	Ур. $\frac{\partial \rho_t}{\partial t} =$	Функционал $\mathcal{F}(\rho) =$
Ур. теплопроводн.	$\Delta \rho$	$\int_{\mathbb{R}^D} \log \frac{d\rho(x)}{dx} d\rho(x)$
Ур. переноса	$\text{div}(\rho \nabla \Phi)$	$\int_{\mathbb{R}^D} \Phi(x) d\rho(x)$
Ур. Фоккера-Планка	$\text{div}(\rho \nabla \Phi) + \frac{1}{\beta}\Delta \rho$	$\int_{\mathbb{R}^D} \Phi(x) d\rho(x) + \frac{1}{\beta} \int_{\mathbb{R}^D} \log \frac{d\rho(x)}{dx} d\rho(x)$





[9] David Alvarez-Melis, Yair Schiff и Youssef Mroueh (2022). “Optimizing Functionals on the Space of Probabilities with Input Convex Neural Networks”. В: Transactions on Machine Learning Research. URL: https://openreview.net/forum?id=dpOYN7o8Jm.

[10] Image source: https://en.wikipedia.org/wiki/Heat_equation

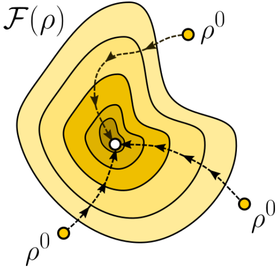
Задача 3: Васерштейн-2 градиентные потоки[11]

Рассмотрим функционал $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^D) \to \mathbb{R}$ на пр-ве вероятностных распределений.

Через $\mathcal{F}'(\rho)$ обозначим <u>первую вариацию</u> функционала $\mathcal{F}$ в точке ρ, т.е. функцию $f_\rho : \mathbb{R}^D \to \mathbb{R}$, для которой

$$\frac{d}{d\epsilon}\mathcal{F}(\rho + \epsilon\chi)|_{\epsilon=0} = \int_{\mathbb{R}^D} f_\rho(x) d\chi(x)$$

для любой меры χ со знаком на $\mathbb{R}^D$ такой, что $\int 1 d\chi(x) = 0$, и существует $\epsilon_0 > 0$, при котором $\rho + \epsilon_0\chi \in \mathcal{P}_2(\mathbb{R}^D)$.



<u>Васерштейн-2 градиентный поток</u> – непрерывная последовательность $\{\rho_t\}_{t\in\mathbb{R}_+}$ в $\mathcal{P}_2(\mathbb{R}^D)$, удовлетворяющая

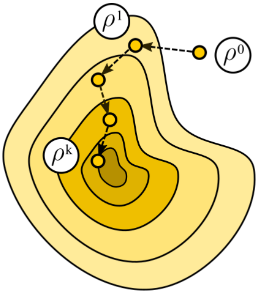
$$\frac{\partial \rho_t}{\partial t} - \text{div}\Big(\rho_t \underbrace{\nabla_x \mathcal{F}'(\rho_t)}_{=\partial_{\mathcal{W}_2}\mathcal{F}(\rho_t)}\Big) = 0 \qquad \text{s.t. } \rho_0 = \rho^0.$$

[11]Filippo Santambrogio (2017). “{Euclidean, metric, and Wasserstein} gradient flows: an overview”. В: Bulletin of Mathematical Sciences 7.1, с. 87—154.

Вычисление Васерштейн-2 градиентного потока[12]

Жордан, Кинделехрер и Отто (JKO) преложили итеративно вычислять дискретную последовательность $\rho^1, \rho^2, \ldots$:

$$\rho^{k+1} \leftarrow \operatorname*{arg\,min}_{\rho \in \mathcal{P}_2(\mathbb{R}^D)} \left[\mathcal{F}(\rho) + \frac{1}{2\tau} \mathcal{W}_2^2(\rho^k, \rho) \right].$$



При $\tau \to 0$ выполняется $\rho^k \approx \rho_{\tau \cdot k}$, т.е. дискретизованный поток ρ^k сходится к истинному потоку ρ_t.

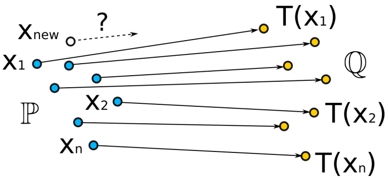
Практическая имплементация схемы JKO нетривиальна, т.к. требует вычисления слагаемого $\mathcal{W}_2$.

Задача 3: Вычисление Васерштейн-2 градиентного потока функционала.

[12]Richard Jordan, David Kinderlehrer и Felix Otto (1998). “The variational formulation of the Fokker–Planck equation”. В: SIAM journal on mathematical analysis 29.1, с. 1—17.

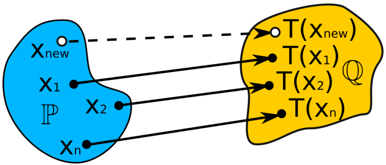
Существующие численные методы OT[13]

Дискретные



+ выпуклая оптимизация;
+ сильные теор. гарантии;
- слабая масштабируемость;
- нет оценок вне выборки;

Параметрические (непрерывные)



± нейронные сети;
± ограниченные теор. гарантии;
+ масштабируемость;
+ оценки вне выборки;

[13]Gabriel Peyré, Marco Cuturi и др. (2019). "Computational optimal transport: With applications to data science". В: Foundations and Trends® in Machine Learning 11.5-6, с. 355—607.

Параметрические (непрерывные) двойственные методы: недостатки

Двойственная постановка Канторовича[14] $(f \in \mathcal{L}^1(\mathbb{P}), g \in \mathcal{L}^1(\mathbb{Q}))$

$$\mathbb{W}_2^2(\mathbb{P}, \mathbb{Q}) = \max_{f,g} \left[\int_{\mathbb{R}^D} f(x) d\mathbb{P}(x) + \int_{\mathbb{R}^D} g(y) d\mathbb{Q}(y) \right] \qquad \text{s.t. } f(x) + g(y) \le \frac{\|x - y\|^2}{2}.$$

Большинство непрерывных методов решают двойственную задачу, т.е., находят f^*, и восстанавливают решение прямой задачи, используя $T^*(x) = x - \nabla f^*(x)$.

1. Регуляризованный OT: наложение мягкого штрафа на f, g за невыполнение условия $f \oplus g \le \frac{1}{2}\| \cdot \|^2$. Данные методы находят <u>смещенное</u> решение задачи;
2. Максиминный OT: оптимизация потенциалов f, g с использованием c-трансформации. Данные методы склонны к <u>численной нестабильности</u>.

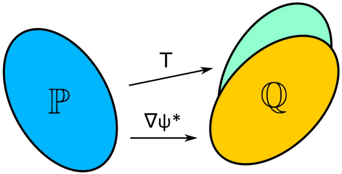
Таким образом, существует необходимость разработки параметрических численных методов, позволяющих находить OT без смещения и максиминной оптимизации.

[14] Cédric Villani (2008). Optimal transport: old and new. T. 338. Springer Science & Business Media.

Задача 4: Построение эталонных пар для Васерштейн-2 ОТ

Вопрос: как оценить качество восстановленного отображения $T : \mathbb{R}^D \to \mathbb{R}^D$ в задаче ОТ?

Проблема: существует ограниченное число пар непрерывных распределений $\mathbb{P}$, $\mathbb{Q}$, для которых аналитически известно истинное оптимальное транспортное отображение $\nabla\psi^*$.



Таким образом, требуется разработка количественной методики тестирования непрерывных методов вычисления ОТ.

Задача 4: Построение эталонных пар распределений для Васерштейн-2 ОТ.

Положения, выносимые на защиту (1 положение = 1 глава работы)

1. Предложен неминимаксный алгоритм для вычисления Васерштейн-2 оптимального отображения и расстояния между непрерывными распределениями, базирующийся на нейронных сетях с выпуклой по входу архитектурой и применении разработанной циклической регуляризации в двойственной задаче. Получены теоретические оценки, связывающие ошибку минимизации разработанного целевого функционала с ошибкой решения целевой задачи нахождения отображения. [прим.: задача 1]
2. Предложен неминимаксный алгоритм для вычисления Васерштейн-2 барицентра семейства непрерывных распределений, базирующийся на нейронных сетях с выпуклой по входу архитектурой и использовании разработанных циклической регуляризации и регуляризации на конгруэнтность двойственных переменных задачи. Получены теоретические оценки, связывающие ошибку минимизации разработанного целевого функционала с ошибкой решения целевой задачи нахождения барицентра. [прим.: задача 2]
3. Предложен численный метод имплементации схемы JKO для моделирования Васерштейн-2 градиентных потоков функционалов на пространствах многомерных вероятностных распределений с помощью использования нейронных сетей выпуклых по входу. [прим.: задача 3]
4. Разработана методология на основе нейронных сетей с выпуклой по входу архитектурой для синтеза пар непрерывных распределений с аналитически известным ОТ отображением между ними. Данная разработка позволяет заполнить имеющийся пробел в тестировании непрерывных методов решения задачи ОТ и исследования математических моделей, на которых они основаны. [прим.: задача 4]

Публикации на основе результатов диссертационной работы

Результаты диссертационной работы опубликованы в 5 научных статьях в трудах 3 конференций [ранг Core A*], из них 2 работы [A1, A2] – индексируемы Scopus.

A1 (Core A*, Scopus) Korotin, A., Li, L., Genevay, A., Solomon, J. M., Filippov, A., & Burnaev, E. (2021). Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark. Advances in Neural Information Processing Systems, 34, 14593-14605.

A2 (Core A*, Scopus) Mokrov, P., Korotin, A., Li, L., Genevay, A., Solomon, J. M., & Burnaev, E. (2021). Large-scale wasserstein gradient flows. Advances in Neural Information Processing Systems, 34, 15243-15256.

A3 (Core A*) Korotin, A., Egiazarian, V., Asadulaev, A., Safin, A., & Burnaev, E. (2020, September). Wasserstein-2 Generative Networks. In International Conference on Learning Representations.

A4 (Core A*) Korotin, A., Li, L., Solomon, J., & Burnaev, E. (2020, September). Continuous Wasserstein-2 Barycenter Estimation without Minimax Optimization. In International Conference on Learning Representations.

A5 (Core A*) Rout, L., Korotin, A., & Burnaev, E. (2021, September). Generative Modeling with Optimal Transport Maps. In International Conference on Learning Representations.

Работы [A1-A5] суммарно имеют 100+ цитирований согласно научной базе Google Scholar.

Презентации результатов диссертации

Презентации на международных конференциях

05.2021 [x2]

International Conference on Learning Representations 2021
ICLR 2021, Core A*, Онлайн
Доклад: Wasserstein-2 Generative Networks
Доклад: Continuous Wasserstein-2 Barycenter Estimation without Minimax Optimization

12.2021 [x2]

Neural Information Processing Systems 2021
NeurIPS 2021, Core A*, Онлайн
Доклад: Large-Scale Wasserstein Gradient Flows
Доклад: Do Neural Optimal Transport Solvers Work? A Continuous Wasserstein-2 Benchmark

05.2022

International Conference on Learning Representations 2021
ICLR 2022, Core A*, Онлайн
Доклад: Generative Modeling with Optimal Transport Maps

Программный код для разработанных моделей публично доступен по ссылке
github.com/iamalexkorotin

Презентации результатов диссертации

Презентации на научных семинарах и воркшопах

10.2019	Huawei Machine Learning Workshop (Sochi, Russia) Доклад: Generative Models [Best presentation award]
06.2020	Seminar of the Geometric Data Processing Group (Онлайн) Доклад: Wasserstein-2 Generative Networks
02.2020	Math of Machine Learning School 2020 (Онлайн) Доклад: Wasserstein-2 Generative Networks
09.2020	SMILES Machine Learning Summer School 2020 (Онлайн) Доклад: Wasserstein-2 Generative Networks
12.2020	Research Seminar on Bayesian Methods in ML (Онлайн) Доклад: Wasserstein-2 Generative Networks

Награды, полученные за научные результаты

Научная премия Яндекса имени Ильи Сегаловича

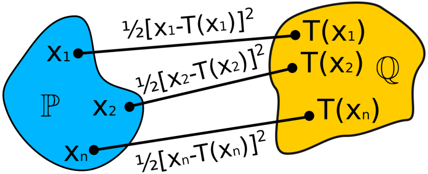
для студентов, проводящих передовые исследования в области машинного обучения.

Лауреат премии в 2019 и 2021 годах

Содержание работы

Задача 1

Вычисление Васерштейн-2 оптимального отображения.



Результат главы

Предложен новый одноэтапный нейросетевой метод для вычисления Васерштейн-2 оптимальных транспортных отображений между непрерывными распределениями без внесения смещения или минимаксной оптимизации.

Предлагаемый неминимаксный функционал для оптимизации

Минимаксный модход (<u>существующее решение</u>)[15]

$$\min_{\psi\in\text{Convex}}\left[\int\psi(x)d\mathbb{P}(x)+\int\overline{\psi}(y)d\mathbb{Q}(y)\right]=$$

$$\min_{\psi\in\text{Conv}}\max_{\phi\in\text{Conv}}\left[\underbrace{\int\psi(x)d\mathbb{P}(x)+\int\Big[\langle\nabla\phi(y),y\rangle-\psi\big(\nabla\phi(y)\big)\Big]d\mathbb{Q}(y)}_{\text{Corr}(\mathbb{P},\mathbb{Q}|\psi,\phi)}\right]$$

<u>Предлагаемый</u> неминимаксный подход: циклический регуляризатор ($\lambda>0$)

$$\min_{\psi,\phi\in\text{Conv}}\text{Corr}(\mathbb{P},\mathbb{Q}\mid\psi,\phi;\lambda)\stackrel{\text{def}}{=}\min_{\psi,\phi\in\text{Conv}}\left[\text{Corr}(\mathbb{P},\mathbb{Q}|\psi,\phi)+\underbrace{\int_{\mathcal{Y}}\frac{\lambda}{2}\|\nabla\psi\big(\nabla\phi(y)\big)-y\|^2d\mathbb{Q}(y)}_{\text{Цикл. рег. }\mathcal{R}_2^{\mathbb{Q}}(\psi,\phi)}\right].$$

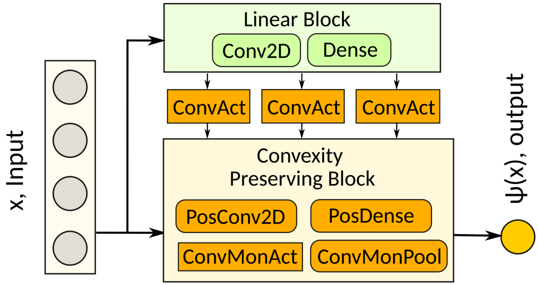
На практике предлагается аппроксимировать $\{\psi,\phi\}$ выпуклыми по входу нейросетями и тренировать их параметры с помощью стохастического град. спуска на выборках из $\mathbb{P},\mathbb{Q}$.

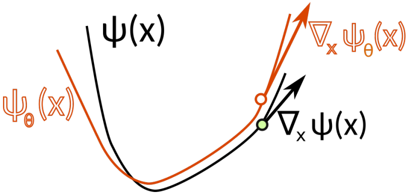
[15] Ashok Makkuva и др. (2020). “Optimal transport mapping via input convex neural networks”. В: International Conference on Machine Learning. PMLR, с. 6672—6681.

Выпуклые по входу нейронные сети (ICNN)

Выпуклые функции можно аппроксимировать нейросетями $\psi(x) : \mathbb{R}^D \to \mathbb{R}$.

- $\psi_\theta : \mathbb{R}^D \to \mathbb{R}$ – глубокая выпуклая по входу нейросеть[16] (ICNN);
- $T_\theta = \nabla_x \psi_\theta : \mathbb{R}^D \to \mathbb{R}^D$ - транспортное отображение.





[16] Brandon Amos, Lei Xu и J Zico Kolter (2017). “Input convex neural networks”. В: Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, с. 146—155.

Теорема 1. Связь решений предложенной двойственной и прямой задач

Пусть $\mathbb{P},\mathbb{Q} \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^{\mathrm{D}})$. Пусть $\psi^* : \mathbb{R}^{\mathrm{D}} \to \mathbb{R}$ – оптимальный выпуклый потенциал:

$$\psi^* = \operatorname*{arg\,min}_{\psi \in \mathrm{Convex}} \mathrm{Corr}(\mathbb{P},\mathbb{Q}|\psi) = \operatorname*{arg\,min}_{\psi \in \mathrm{Convex}} \Big[\int_{\mathbb{R}^{\mathrm{D}}} \psi(x) d\mathbb{P}(x) + \int_{\mathbb{R}^{\mathrm{D}}} \overline{\psi}(y) d\mathbb{Q}(y) \Big].$$

Пусть дифф. выпуклые функции $\hat{\psi}, \hat{\phi} : \mathbb{R}^{\mathrm{D}} \to \mathbb{R}$ таковы, что, при некотором $\epsilon \in \mathbb{R}$,

$$\mathrm{Corr}\Big(\mathbb{P},\mathbb{Q} \mid \hat{\psi},\hat{\phi};\lambda\Big) \leq \Big[\int_{\mathbb{R}^{\mathrm{D}}} \psi^*(x) d\mathbb{P}(x) + \int_{\mathbb{R}^{\mathrm{D}}} \overline{\psi^*}(y) d\mathbb{Q}(y) \Big] + \epsilon = \mathrm{Corr}(\mathbb{P},\mathbb{Q} \mid \psi^*) + \epsilon. \tag{1}$$

Предположим, что $\hat{\psi}$ является β-сильно выпуклой ($\beta > \frac{1}{\lambda} > 0$) и $\mathcal{B}$-гладкой ($\mathcal{B} \geq \beta$). Предположим также, что $\hat{\phi}$ имеет биективный градиент $\nabla\hat{\phi}$. Тогда выполнено:

1. Оценка сверху для корреляции

$$\mathrm{Corr}\Big(\mathbb{P},\mathbb{Q} \mid \hat{\psi},\hat{\phi};\lambda\Big) \geq \mathrm{Corr}\Big(\mathbb{P},\mathbb{Q} \mid \psi^*\Big) \qquad (\text{т.е. } \epsilon \geq 0);$$

2. Прямое и обратное транспортные свойства

$$\frac{1}{2}\int_{\mathbb{R}^{\mathrm{D}}} \|\nabla\hat{\psi}(x) - \nabla\psi^*(x)\|^2 d\mathbb{P}(x) \leq \frac{(\mathcal{B})^2 \cdot \epsilon}{\lambda\beta - 1} \cdot \Big[\frac{1}{\sqrt{\beta}} + \sqrt{\lambda}\Big]^2 = O(\epsilon);$$

$$\frac{1}{2}\int_{\mathbb{R}^{\mathrm{D}}} \|\nabla\hat{\phi}(y) - \nabla\overline{\psi^*}(y)\|^2 d\mathbb{Q}(y) \leq \frac{\epsilon}{\beta - \frac{1}{\lambda}} = O(\epsilon).$$

Теорема 2. Свойство решений в ограниченном классе функций

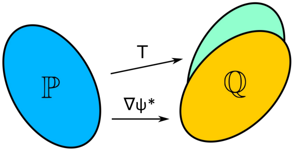
Пусть $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^{\mathrm{D}})$. Пусть $\psi^* : \mathbb{R}^{\mathrm{D}} \to \mathbb{R}$ – оптимальный выпуклый потенциал. Далее, пусть $\boldsymbol{\Psi}, \boldsymbol{\Phi}$ – классы дифференцируемых выпуклых функций $\mathbb{R}^{\mathrm{D}} \to \mathbb{R}$ соответственно. Предположим, что

1. $\exists \psi_1 \in \boldsymbol{\Psi}$ с ϵ_1-близким градиентом прямого отображения $\nabla \psi^*$ в смысле $\mathcal{L}^2(\mathbb{R}^{\mathrm{D}} \to \mathbb{R}^{\mathrm{D}}, \mathbb{P})$:
$$\|\nabla \psi_1 - \nabla \psi^*\|^2_{\mathbb{P}} \stackrel{\text{def}}{=} \int_{\mathbb{R}^{\mathrm{D}}} \|\nabla \psi_2(\mathrm{y}) - \nabla \psi^*(\mathrm{y})\|^2 d\mathbb{P}(\mathrm{y}) \leq \epsilon_1,$$
где ψ_1 является $\mathcal{B}$-гладкой;
2. $\exists \phi_2 \in \boldsymbol{\Phi}$ с ϵ_2-близким градиентом к обратному отображению $\nabla \overline{\psi^*}$ в смысле $\mathcal{L}^2(\mathbb{R}^{\mathrm{D}} \to \mathbb{R}^{\mathrm{D}}, \mathbb{Q})$:
$$\|\nabla \phi_2 - \nabla \overline{\psi^*}\|^2_{\mathbb{Q}} \stackrel{\text{def}}{=} \int_{\mathbb{R}^{\mathrm{D}}} \|\nabla \phi_2(\mathrm{y}) - \nabla \overline{\psi^*}(\mathrm{y})\|^2 d\mathbb{Q}(\mathrm{y}) \leq \epsilon_2.$$

Пусть $(\hat{\psi}, \hat{\phi})$ – минимизаторы рег. корреляций в $\boldsymbol{\Psi} \times \boldsymbol{\Phi}$: $(\hat{\psi}, \hat{\phi}) = \arg\min_{\psi \in \boldsymbol{\Psi}, \psi \in \boldsymbol{\Phi}} \mathrm{Corr}\big(\mathbb{P}, \mathbb{Q} \mid \psi, \phi; \lambda\big)$.
Тогда регуляризованные корреляции для $(\hat{\psi}, \hat{\phi})$ удовлетворяют следующему неравенству:
$$\mathrm{Corr}\big(\mathbb{P}, \mathbb{Q} \mid \hat{\psi}, \hat{\phi}; \lambda\big) \leq \mathrm{Corr}(\mathbb{P}, \mathbb{Q}) + \Big[\frac{\lambda}{2}(\mathcal{B}\sqrt{\epsilon_2} + \sqrt{\epsilon_1})^2 + (\mathcal{B}\sqrt{\epsilon_2} + \sqrt{\epsilon_1}) \cdot (\sqrt{\epsilon_2}) + \frac{\mathcal{B}}{2}\epsilon_2\Big],$$
т.е. регуляризованные корреляции не превышают истинных плюс $\mathrm{O}(\epsilon_1 + \epsilon_2)$.

Эксперименты: ОТ между гауссианами (количественные результаты)

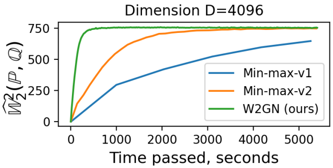
Гауссовский случай: $\mathbb{P}, \mathbb{Q} = \mathcal{N}(\mu_{\mathbb{P}}, \Sigma_{\mathbb{P}}), \mathcal{N}(\mu_{\mathbb{Q}}, \Sigma_{\mathbb{Q}})$



Metric:

$$\mathcal{L}^2\text{-UVP}(\hat{\mathrm{T}}) = 100 \cdot \frac{\|\hat{\mathrm{T}} - \nabla\psi^*\|^2_{\mathbb{P}}}{\mathrm{Var}(\mathbb{Q})}\%$$

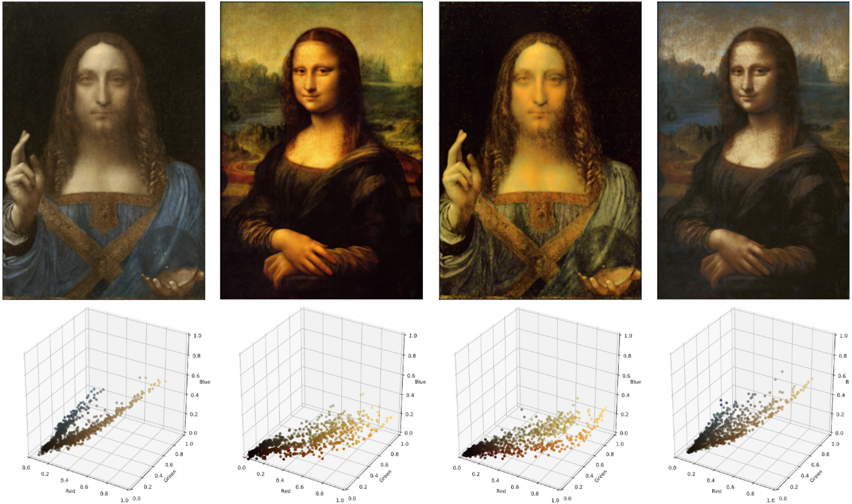
Размерность	2	4	8	16	32	64	128	256	512	1024	2048	4096
LSOT	<1	3.7	7.5	14.3	23	34.7	46.9	>50				
MM-1	<1	<1	<1	<1	<1	1.2	1.4	1.3	1.5	1.6	1.8	2.7
MM-2	<1	<1	<1	<1	<1	<1	1	1.1	1.2	1.3	1.5	2.1
W2GN (ours)	<1	<1	<1	<1	<1	<1	1	1.1	1.3	1.3	1.8	1.5



Согласно экспериментальным результатам, предложенный метод работает лучше/сравнимо существующих методов, но эмпирически быстрее сходится.

Передача цвета: качественные результаты

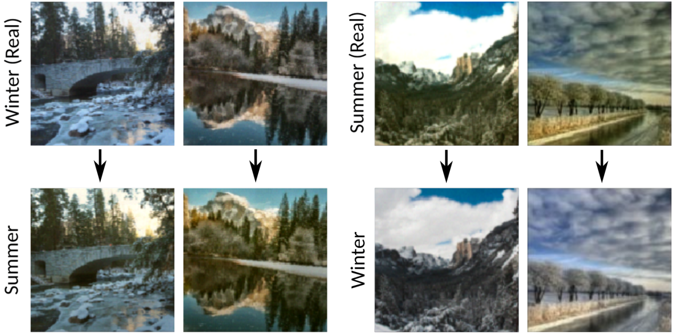
Входные изображения → выходные изображения



Непарный перенос стиля между наборами изображений: качественные рез-ты

Изображения 128×128

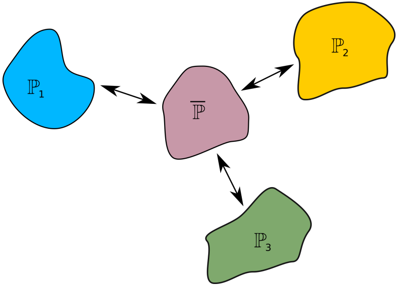
Набор изображений Winter2SummerYosemite[17]



[17]Jun-Yan Zhu и др. (2017). “Unpaired image-to-image translation using cycle-consistent adversarial networks”. В: Proceedings of the IEEE international conference on computer vision, с. 2223—2232.

Задача 2

Вычисление Васерштейн-2 барицентра.



Результат главы

Предложен нейросетевой метод вычисления барицентров Васерштейна-2 между непрерывными распределениями, основанный на выпуклых по входу нейросетях сетях и неминимаксмной оптимизации без смещения.

Прямая и двойственная задачи Васерштейн-2 барицентра[18]

Прямая постановка:

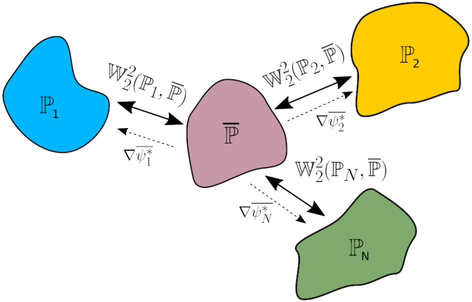
$$\overline{\mathbb{P}} = \arg\min_{\mathbb{P}\in\mathcal{P}_2(\mathbb{R}^D)} \underbrace{\sum_{n=1}^{N} \alpha_n \mathbb{W}_2^2(\mathbb{P}, \mathbb{P}_n)}_{\mathcal{B}(\mathbb{P})}.$$

Двойственная постановка:

$$\mathcal{B}(\overline{\mathbb{P}}) = \text{Const} - \min_{\{\psi_n\}\text{ congr.}} \underbrace{\sum_{n=1}^{N} \alpha_n \int_{\mathbb{R}^D} \psi_n(y) d\mathbb{P}_n(y)}_{\text{MultiCorr}(\{\alpha_n,\mathbb{P}_n\}|\{\psi_n\})},$$

где выпуклые ψ_n согласованы (конгруэнтны):

$$\forall x \in \mathbb{R}^D : \qquad \sum_{n=1}^{N} \alpha_n \overline{\psi_n}(x) = \frac{\|x\|^2}{2}.$$



Прямая-двойственная связь:

$$\nabla\psi_n^* \sharp \mathbb{P}_n = \overline{\mathbb{P}}$$

позволяет восстановить $\overline{\mathbb{P}}$

из двойственных потенциалов ψ_n^*.

[18] Martial Agueh и Guillaume Carlier (2011). “Barycenters in the Wasserstein space”. В: SIAM Journal on Mathematical Analysis 43.2, с. 904—924.

Предложенный неминимаксный функционал оптимизации

Рассмотрим оптимизацию ($\lambda > 0, \tau \geq 1$) по 2N выпуклым функциям $\{\psi_n, \phi_n\}$:

$$\min_{\{\psi_n,\phi_n\}} \overbrace{\sum_{n=1}^{N}\left[\alpha_n \int_{\mathbb{R}^D}[\langle x, \nabla\psi_n(x)\rangle - \phi_n(\nabla\psi_n(x))]d\mathbb{P}_n(x)\right]}^{\text{Аппроксимация множественных корреляций}} + \underbrace{\tau \cdot \mathcal{R}_1^{\widehat{\mathbb{P}}}(\{\phi_n\})}_{\text{Рег. на согл.}} + \underbrace{\lambda\sum_{n=1}^{N}\alpha_n \mathcal{R}_2^{\mathbb{P}_n}(\psi_n, \phi_n)}_{\text{Цикл. рег.}}.$$

Здесь $\mathcal{R}_2^{\mathbb{P}_n}(\psi_n, \phi_n)$ – циклические регуляризаторы, а $\mathcal{R}_1^{\widehat{\mathbb{P}}}(\{\phi_n\})$ – новый регуляризатор на согласованность (с вспомогательным распределением $\hat{\mathbb{P}}$):

$$\mathcal{R}_1^{\widehat{\mathbb{P}}}(\{\phi_n\}) = \int_{\mathbb{R}^D}\Big[\sum_{n=1}^{N}\alpha_n\phi_n(y) - \frac{\|y\|^2}{2}\Big]_+ d\widehat{\mathbb{P}}.$$

На практике предлагается аппроксимировать функции $\{\psi_n, \phi_n\}$ выпуклыми по входу нейронными сетями и тренировать их параметры с помощью стохастического градиентного спуска, используя случайные выборки из $\mathbb{P}_n, \hat{\mathbb{P}}$.

Предложенное теоретическое обоснование

Теорема 3. Связь решений двойственной и прямой задач

Пусть $\overline{\mathbb{P}} \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^{\mathrm{D}})$ – барицентр для $\mathbb{P}_1, \dots, \mathbb{P}_{\mathrm{N}} \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^{\mathrm{D}})$ с весами $\alpha_1, \dots, \alpha_{\mathrm{N}}$ и пусть $\{\psi_{\mathrm{n}}^*\}$ – оптимальные конгруэнтные потенциалы для задачи о барицентре. Предположим, что $\tau, \hat{\mathbb{P}}$ таковы, что $\tau \geq 1$ и $\tau \cdot \widehat{\mathbb{P}} \geq \overline{\mathbb{P}}$. Пусть нам также даны выпуклые потенциалы $\{\hat{\psi}_{\mathrm{n}}\}$ и β-сильно выпуклые и $\mathcal{B}$-гладкие выпуклые потенциалы $\{\hat{\phi}_{\mathrm{n}}\}$ с $0 < \beta \leq \mathcal{B} < \infty$ и $\lambda > \frac{\mathcal{B}}{2\beta^2}$. Положим

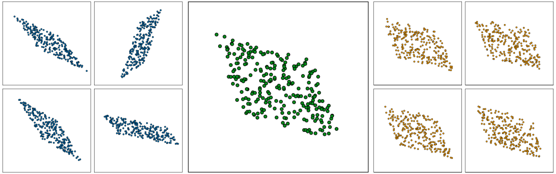
$$\epsilon = \mathrm{MultiCorr}\big(\{\alpha_{\mathrm{n}}, \mathbb{P}_{\mathrm{n}}\} \mid \{\hat{\psi}_{\mathrm{n}}\}, \{\hat{\phi}_{\mathrm{n}}\}; \tau, \widehat{\mathbb{P}}, \lambda\big) - \mathrm{MultiCorr}\big(\{\alpha_{\mathrm{n}}, \mathbb{P}_{\mathrm{n}}\} \mid \{\psi_{\mathrm{n}}^*\}\big).$$

Тогда имеем $\epsilon \geq 0$ и при всех $\mathrm{n} \in \{1, \dots, \mathrm{N}\}$

$$\int_{\mathbb{R}^{\mathrm{D}}} \|\nabla\hat{\psi}_{\mathrm{n}}(\mathrm{y}) - \nabla\psi^*(\mathrm{y})\|^2 d\mathbb{P}_{\mathrm{n}}(\mathrm{y}) \leq \frac{2\epsilon}{\alpha_{\mathrm{n}}} \cdot \left(\sqrt{\frac{1}{\beta}} + \sqrt{\frac{1}{\lambda\beta^2 - \frac{\mathcal{B}}{2}}}\right)^2 = O(\epsilon), \tag{2}$$

что, в частности, означает $\mathbb{W}_2^2(\nabla\hat{\psi}_{\mathrm{n}}\sharp\mathbb{P}_{\mathrm{n}}, \overline{\mathbb{P}})$.

Количественное сравнение в линейно-разбросанном случае[19]



Метрика	Метод	D=2	4	8	16	32	64	128	256
$\mathrm{B}\mathbb{W}_2^2$-UVP, %	[FC$\mathbb{W}$B], Cuturi и Doucet 2014	0.64	0.77	1.22	3.75	8.92	14.3	18.46	21.64
	[SC$\mathbb{W}_2$B], Fan, Taghvaei и Chen 2021	0.12	0.10	0.19	0.29	0.46	0.6	1.38	2.9
$\mathcal{L}_2$-UVP, % (potentials)		0.17	0.12	0.2	0.31	0.47	0.62	1.21	1.52
	[CR$\mathbb{W}$B], L. Li и др. 2020	0.58	1.83	8.09	21.23	55.17	>100		
	[C$\mathbb{W}_2$B], ours	0.17	0.08	0.06	0.1	0.2	0.25	0.42	0.82

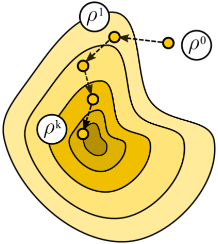
Таблица 1: Сравнение метрики UVP в линейно-разбросанном случае (cube uniform), $N = 4$.

Согласно экспериментальным результатам, предложенный метод превосходит по качеству существующие методы.

[19] Pedro C Álvarez-Esteban и др. (2016). “A fixed-point approach to barycenters in Wasserstein space”. В: Journal of Mathematical Analysis and Applications 441.2, с. 744—762.

Задача 3

Вычисление Васерштейн-2 градиентного потока.



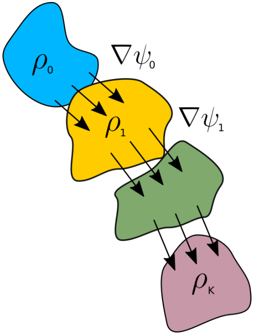
Результат главы

Предложен численный метод имплементации схемы JKO для моделирования Васерштейн-2 градиентных потоков функционалов на пространствах многомерных вероятностных распределений с помощью использования нейронных сетей.

Предлагаемый алгоритм вычисления Васерштейн-2 градиентного потока

Напомним шаг JKO схемы:

$$\rho^{k+1} \leftarrow \operatorname*{arg\,min}_{\rho \in \mathcal{P}_2(\mathbb{R}^D)} \left[\mathcal{F}(\rho) + \frac{1}{2\tau}\mathcal{W}_2^2(\rho^k, \rho)\right].$$



В диссертации предлагается заменить оптимизацию по распределениям на эквивалентную оптимизацию по выпуклым функциям ψ:

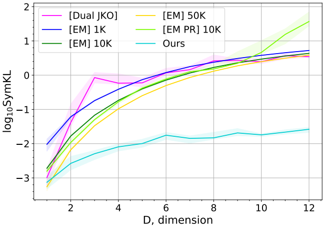
$$\psi_k \leftarrow \operatorname*{arg\,min}_{\psi \in \mathrm{Conv}} \left[\mathcal{F}(\nabla\psi \sharp \rho^k) + \frac{1}{\tau}\int_{\mathbb{R}^D} \frac{1}{2}\|x - \nabla\psi(x)\|^2 d\rho^k(x)\right].$$

При данном подходе $\rho^k = \nabla\psi_{k-1}\sharp\left[\nabla\psi_{k-2}\sharp\left[\dots\nabla\psi_0\sharp\rho^0\right]\right]$.

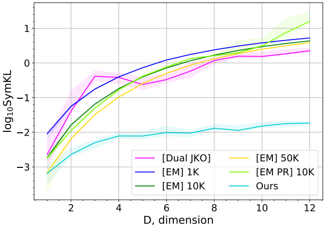
Для численного осуществления шагов JKO предлагается обучать последовательность выпуклых по входу нейронных сетей $\{\psi_k\}$ и оптимизировать их параметры с помощью стохастического градиентного спуска, используя случайные выборки из ρ^k.

Количественное сравнение в случае процессов Орнштейна—Уленбека

При $\Phi(x) = \frac{1}{2}(x - b)^T A(x - b)$ с $A \succeq 0$ и гауссовским ρ^0 град. поток функ-ла $\mathcal{F}_{FP}(\rho) = \int_{\mathbb{R}^D} \Phi(x) d\rho(x) + \frac{1}{\beta} \int_{\mathbb{R}^D} \log \frac{d\rho(x)}{dx} d\rho(x)$ аналитически вычислим для всех $t \geq 0$.[20]



$t = 0.5$



$t = 0.8$

Согласно экспериментальным результатам, предложенный метод превосходит по качеству существующие методы (для сравнения используется метрика SymKL).

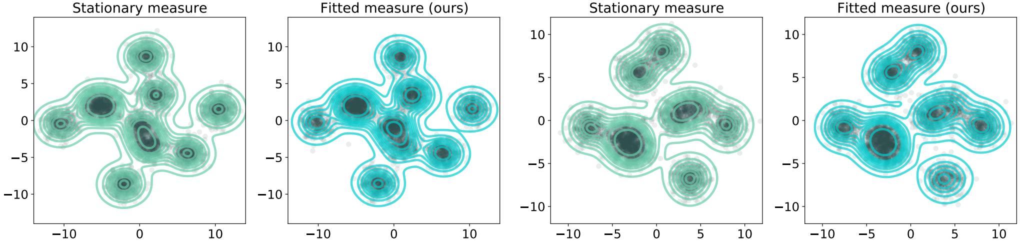
[20] Pat Vatiwutipong и Nattakorn Phewchean (2019). “Alternative way to derive the distribution ofthe multivariate Ornstein–Uhlenbeck process”. В: Advances in Difference Equations 276.

Качественное сравнение: сходимость к стационарному распределению

Стартуя из начальной точки ρ^0, градиентный поток $\mathcal{F}_{\mathrm{FP}}$ сходится к[21]

$$\frac{\mathrm{d}\rho^*}{\mathrm{dx}}(\mathrm{x}) = \mathrm{Z}^{-1}\exp(-\beta\Phi(\mathrm{x})),$$

где $\mathrm{Z} = \int_{\mathbb{R}^\mathrm{D}} \exp(-\beta\Phi(\mathrm{x}))\mathrm{dx}$ – нормировочная константа.

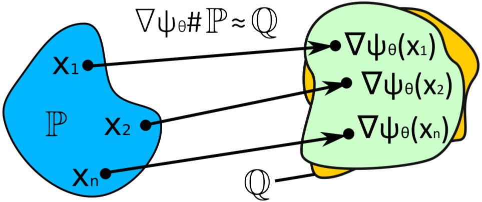


2 главные компоненты стационарного и обученного распределений, D = 13 (слева) и D = 32 (справа).

[21]Hannes. Risken (1996). The Fokker-Planck Equation: Methods of Solution and Applications (Springer Series in Synergetics). Springer,

Задача 4

Построение эталонных пар для Васерштейн-2 транспорта.

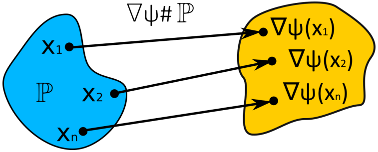


Результат главы

Предложена общая методология построения пар непрерывных распределений с аналитически известными ОТ отображениями для квадратичной стоимости для тестирования параметрических методов ОТ.

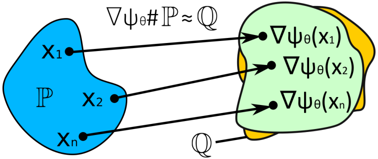
Ключевая предложенная идея построения эталонных пар

Для $\mathbb{P} \in \mathcal{P}_{2,\text{ac}}(\mathbb{R}^D)$ пусть $\nabla\psi\sharp\mathbb{P}$ будет его образом под действием градиента выпуклой функции $\psi : \mathbb{R}^D \to \mathbb{R}$.



Тогда $\nabla\psi$ является ОТ отображением между $\mathbb{P}$ и $\nabla\psi\sharp\mathbb{P}$.

Предлагается приближать ОТ отображение с помощью градиента выпуклой по входу нейросети $\nabla\psi_\theta$



и затем использовать пару $(\mathbb{P}, \nabla\psi_\theta\sharp\mathbb{P}) \approx (\mathbb{P}, \mathbb{Q})$ как эталонную с известным ОТ отображением $\nabla\psi_\theta$.

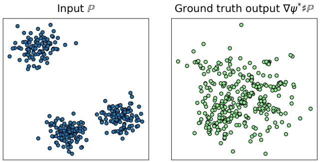
В основе предложенной идеи лежит теорема Бренье[22].

[22] Yann Brenier (1991). “Polar factorization and monotone rearrangement of vector-valued functions”. В: Communications on pure and applied mathematics 44.4, с. 375—417.

Разработанные эталонные пары вероятностных распределений

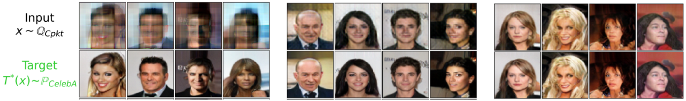
Многомерные пары распределений

$\mathrm{D} = 2, 4, 8, \ldots, 256$



Пары распределений в пространстве изображений (на основе Celeba[23])

3 пары, $\mathrm{D} = 12288$



[23] Ziwei Liu и др. (дек. 2015). "Deep Learning Face Attributes in the Wild". В: Proceedings of International Conference on Computer Vision (ICCV).

Количественные результаты метода на разработанных эталонных парах

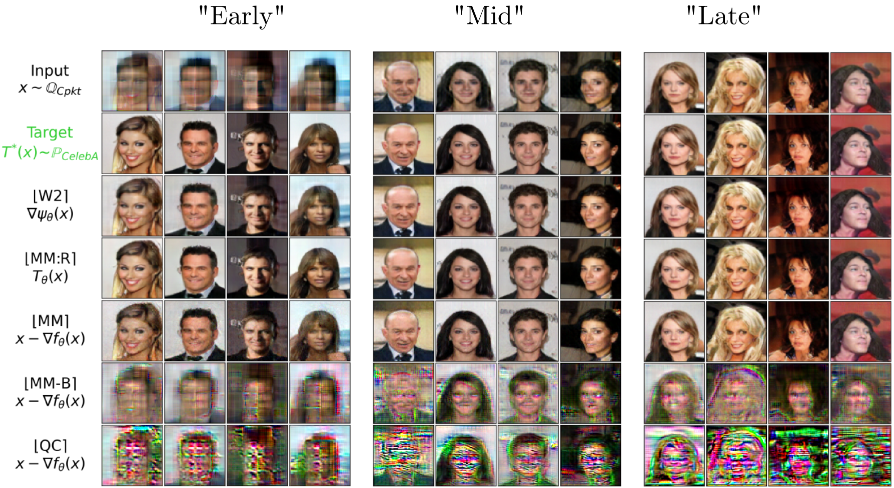
Dim	2	4	8	16	32	64	128	256
⌊W2⌉ (Ours)	0.1	0.7	2.6	3.3	6.0	7.2	2.0	2.7
⌊W2:R⌉	0.2	0.9	4.0	5.3	5.2	7.0	2.0	2.7
⌊MMv1⌉	0.2	1.0	1.8	1.4	6.9	8.1	2.2	2.6
⌊MM⌉	0.1	0.3	0.9	2.2	4.2	3.2	3.1 $\looparrowright$	4.1 $\looparrowright$
⌊MM:R⌉	0.1	0.3	0.7	1.9	2.8	4.5	$\nrightarrow$	$\nrightarrow$
⌊MMv2⌉	0.1	0.68	2.2	3.1	5.3	10.1 $\looparrowright$	3.2 $\looparrowright$	2.7 $\looparrowright$
⌊MMv2:R⌉	0.1	0.7	4.4	7.7	5.8	6.8	2.1	2.8
⌊MM-B⌉	0.1	0.7	3.1	6.4	12.0	13.9	19.0	22.5
⌊LS⌉	5.0	11.6	21.5	31.7	42.1	40.1	46.8	54.7
⌊L⌉	14.1	14.9	27.3	41.6	55.3	63.9	63.6	67.4
⌊QC⌉	1.5	14.5	28.6	47.2	64.0	75.2	80.5	88.2
⌊C⌉	100	100	100	100	100	100	100	100
⌊ID⌉	32.7	42.0	58.6	87	121	137	145	153

Размерности $D = 2, 2^2, \dots, 2^8$. Оранжевый цвет показывает $\mathcal{L}^2$-UVP $> 10\%$.
Красный цвет выделяет значения метрики ниже, чем у простого линейного ⌊L⌉ бейзлайна.

Экспериментальные результаты подтверждают, что

1. Седловые методы ⌊MM⌉ численно нестабильны и могут расходиться ($\looparrowright$);
2. Регуляризованный метод ⌊LS⌉ находит смещенное решение;
3. Предложенный метод ⌊W2⌉ стабилен и работает сравнимо/лучше альтернатив.

Эталонные пары изображений: качественные результаты

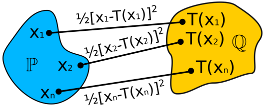


Заключение

Итог: 4 задачи OT = 4 решения = 4 положения

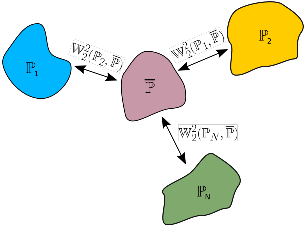
[Задача 1]

Вычисление $\mathbb{W}_2$ OT отображений



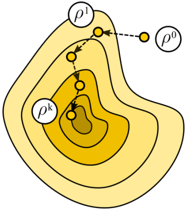
[Задача 2]

Вычисление $\mathbb{W}_2$ барицентров



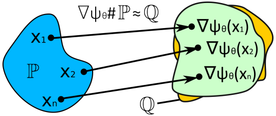
[Задача 3]

Вычисление $\mathbb{W}_2$ градиентных потоков



[Задача 4]

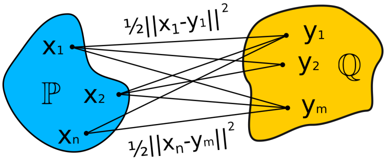
Построение эталонных пар для $\mathbb{W}_2$



Оптимальный транспорт с квадратичной ценой[24] (по Канторовичу)

Квадрат расстояния Васерштейн-2 между $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^D)$ определяется как

$$\mathbb{W}_2^2(\mathbb{P}, \mathbb{Q}) \stackrel{\text{def}}{=} \min_{\pi \in \Pi(\mathbb{P}, \mathbb{Q})} \int_{\mathbb{R}^D \times \mathbb{R}^D} \frac{1}{2} \|x - y\|^2 d\pi(x, y),$$



где минимум берется по всем распределениям π на $\mathbb{R}^D \times \mathbb{R}^D$,
чьи проекции на первую и вторую компоненту равны $\mathbb{P}$ и $\mathbb{Q}$ соответсвенно.

Распределение $\pi^* \in \Pi(\mathbb{P}, \mathbb{Q})$, доставляющее min, называется ОТ планом.

[24] Cédric Villani (2008). Optimal transport: old and new. T. 338. Springer Science & Business Media.

Уравнение непрерывности[25] (Секц. 8.1)

Рассмотрим поле скоростей $\mathrm{v} : (\mathrm{x}, \mathrm{t}) \mapsto \mathrm{v_t(x)} \in \mathbb{R}^{\mathrm{D}}$, для которого выполнено

$$\int_0^{\mathrm{T}} \int_{\mathbb{R}^{\mathrm{D}}} |\mathrm{v_t(x)}| \mathrm{d}\mu_{\mathrm{t}}(\mathrm{x}) \mathrm{dt} < \infty.$$

Последовательность распределений $\{\rho_{\mathrm{t}}\}_{\mathrm{t} \in (0, \mathrm{T})}$ является решением уравнения непрерывности

$$\frac{\partial \rho_{\mathrm{t}}}{\partial \mathrm{t}} + \mathrm{div}(\rho_{\mathrm{t}} \mathrm{v_t}) = 0,$$

если для любой $\mathrm{h} \in \mathrm{C}_{\mathrm{c}}^{\infty}(\mathbb{R}^{\mathrm{D}})$ выполняется равенство

$$\int_0^{\mathrm{T}} \int_{\mathbb{R}^{\mathrm{D}}} \left(\frac{\partial}{\partial \mathrm{t}} \mathrm{h}(\mathrm{x}, \mathrm{t}) + \langle \mathrm{v_t(x)}, \nabla_{\mathrm{x}} \mathrm{h}(\mathrm{x}, \mathrm{t}) \rangle \right) \mathrm{d}\rho_{\mathrm{t}}(\mathrm{x}) \mathrm{dt} = 0.$$

[25] Luigi Ambrosio, Nicola Gigli и Giuseppe Savaré (2005). Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media.

Абсолютно непрерывная кривая[26]

Кривая распределений $\{\rho_t\}_{t\in[0,T]}$ в $\big(\mathcal{P}_2(\mathbb{R}^D), \mathcal{W}_2\big)$ называется абсолютно непрерывной, если существует $m \in \mathcal{L}^1([0,T])$ такое, что для любых $0 < s \le t < T$ выполнено

$$\mathcal{W}_2\big(\rho_s, \rho_t\big) \le \int_s^t m(r)dr.$$

У такой кривой для почти всех $t \in [0,T]$ существует предел

$$\lim_{s\to t} \frac{\mathcal{W}_2(\rho_t, \rho_s)}{|t-s|} = |\rho'|(t),$$

который называется метрической производной. Для почти всех t верно $|\rho'|(t) \le m(t)$.

Если γ - абсолютно непрерывная (и спрямляемая), то $L\big(\rho, [0,T]\big) = \int_0^T |\rho'|(t)dt$, где

$$L\big(\rho, [0,T]\big) = \sup_{0=s_1\le s_2\le\dots\le s_n=T} \sum_{i=1}^{n-1} \mathcal{W}_2\big(\rho_{s_i}, \rho_{s_{i+1}}\big) \text{ - длина кривой.}$$

[26]Полагаем $\mathcal{W}_2 = \sqrt{2}\mathbb{W}_2$.

Связь абсолютно непрерывных кривых и уравнения непрерывности

Теорема 8.3.1.[27] Пусть $\mathrm{I} \subset \mathbb{R}$ - открытый интервал и $\rho_t : \mathrm{I} \to \mathcal{P}_2(\mathbb{R}^D)$ – абсолютно непрерывная кривая. Тогда существует векторное поле $v : (x, t) \mapsto v_t(x)$, для которого

$$v_t \in \mathcal{L}^2(\rho_t) \qquad \text{и} \qquad \|v_t\|_{\mathcal{L}^2(\rho_t)} = |\rho'|(t) \text{ для п.в. } t \in \mathrm{I},$$

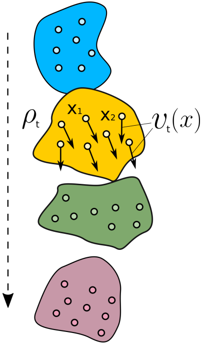
и выполняется уравнение непрерывности

$$\frac{\partial \rho_t}{\partial t} + \mathrm{div}(\rho_t v_t) = 0.$$

При этом v_t называется касательным полем к кривой ρ_t и выполнено

$$v_t \in \overline{\{\nabla\phi : \phi \in C_c^\infty\}}_{\mathcal{L}^2(\rho_t)} = \mathrm{Tan}_{\rho_t}\mathcal{P}_2(\mathbb{R}^D),$$

где последнее называется касательным пространством к ρ_t в $\mathcal{P}_2(\mathbb{R}^D)$.



[27] Luigi Ambrosio, Nicola Gigli и Giuseppe Savaré (2005). Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media.

Дифференцирование на пространстве распределений[28]

В Евклидовом пространстве $\mathbb{R}^D$	В пространстве распределений $\big(\mathcal{P}_2(\mathbb{R}^D), \mathcal{W}_2\big)$
$\mathrm{f} : \mathbb{R}^D \to \mathbb{R}$	$\mathcal{F} : \mathcal{P}_2(\mathbb{R}^D) \to \mathbb{R}$
$\mathrm{x} \in \mathbb{R}^D$	$\mu \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^D)$
$\mathrm{f}(\mathrm{y}) - \mathrm{f}(\mathrm{x}) =$ $\langle \xi, \mathrm{y} - \mathrm{x} \rangle + o(\lVert \mathrm{y} - \mathrm{x} \rVert)$	$\mathcal{F}(\nu) - \mathcal{F}(\mu) =$ $\int_{\mathbb{R}^D} \langle \xi(\mathrm{x}), \mathrm{T}_{\mu\to\nu}(\mathrm{x}) - \mathrm{x} \rangle d\mu(\mathrm{x}) + o(\mathcal{W}_2(\mu, \nu))$
Производная по Фреше $\xi \in \partial \mathrm{f}(\mathrm{x}) \subset \mathbb{R}^D$	$\mathcal{W}_2$-производная (векторное поле $\mathbb{R}^D \to \mathbb{R}^D$) $\xi \in \partial_{\mathcal{W}_2} \mathcal{F}(\rho) \subset \mathrm{Tan}_\mu \mathcal{P}_2(\mathbb{R}^D)$

[28] Для простоты полагаем, что μ абсолютно непрерывно.

Градиентный поток на пространстве распределений[29]

В Евклидовом пространстве $\mathbb{R}^D$	В пространстве распределений $\big(\mathcal{P}_2(\mathbb{R}^D), \mathcal{W}_2\big)$
$f : \mathbb{R}^D \to \mathbb{R}$	$\mathcal{F} : \mathcal{P}_2(\mathbb{R}^D) \to \mathbb{R}$
$x : [0, T] \to \mathbb{R}^D$	$\rho : [0, T] \to \big(\mathcal{P}_2(\mathbb{R}^D), \mathcal{W}_2\big)$
$\begin{cases} \dot{x}(t) = v_t \\ v_t \in -\partial f\big(x(t)\big) \\ x(0) = x^0 \end{cases}$	$\begin{cases} v_t \text{ - кас. поле к } \rho_t \\ v_t \in -\partial_{\mathcal{W}_2} \mathcal{F}\big(\rho_t\big) \\ \rho_0 = \rho^0 \end{cases}$
$\dot{x}(t) = f'\big(x(t)\big),$ $x(0) = x^0$	$\frac{\partial \rho_t}{\partial t} - \mathrm{div}\big(\rho_t \nabla_x \mathcal{F}'(\rho_t)\big) = 0,$ $\rho_0 = \rho^0$

[29] В правом столбце используется связь первой вариации и $\mathcal{W}_2$-производной: $\partial_{\mathcal{W}_2} \mathcal{F}(\rho) = \nabla_x \mathcal{F}'(\rho)$.