DomainBed

起因

当前的机器学习系统在面对新的示例分布时会反复无常地失败，阻碍了这项技术不能在关键领域的使用.

我们希望找到鲁棒的机器学习模型，能够规避假相关。让模型发现不因域内域外而改变的模式

结论

- ERM (Empirical risk minimization) Baseline优于其他结果，主要原因有：
    - 更大的网络架构（ResNet-50）
    - 强大的数据增强
    - 仔细的参数调优
- 强数据增强可以提高分布外泛化，同时不影响分布内泛化
- 消除虚假相关性

评价

各个数据集上的精度不加权重的取均值不太合理。导致所有模型的结果都在66%左右。

一个有趣的现象是，所有模型在CMNIST上的结果都不高，50%左右。据我了解这是份加了颜色的手写字数据集，不是很复杂，精度应该比较高才对。值得深究的一个点。

further

CMNIST值得看一下

此论文中实验部分有必要详细看一下，看下他怎么拿数据集进行训练的。以至于CMNIST上精度那么低。

Risk minimization

Paper title: Principles of Risk Minimization

for Learning Theory
Learning is posed as a problem of function estimation, for which two principles of solution are considered: **empirical risk minimization** and **structural risk minimization**. These two principles are applied to two different statements of the function estimation problem: **global and local**. Systematic improvements in prediction power are illustrated in application to zip-code recognition.

AE-OT generator

值得注意的点

- DNNs 只能表示连续映射
- 生成器是传输映射，传输白噪音因分布到数据集分布. 通常这种映射是不连续的.
- We propose that these phenomena relates deeply with the singularities of distribution transport maps. （我们认为这些现象与分布输运图映射的奇异性密切相关。）
- 流形分布假设在深度学习中被广泛接受，它假设特定类别的自然数据的分布集中在嵌入在高维数据空间中的低维流形上。
- 因此gan和vae隐含的目标是完成两个主要任务:
    - 流形嵌入:寻找嵌入到图像空间的数据流形与潜在空间之间的编码/解码映射;（to find the encoding/decoding maps between the data manifold embedded in the image space and the latent space;）
    - 概率分布传输:将给定的白噪声分布传输到数据分布中，可以传输到潜在空间中，也可以传输到图像空间中。probability distribution transport: to transport a given white noise distribution to the data distribution, either in the latent or in the image space.

做法

将映射拆成两部分：流形嵌入+最优传输.

理论

- 模式坍塌原因，概括一下，目标集**非凸**.
    - 问题：非凸会带来什么影响？凸集又有哪些优势？
- (single mode)
    - Brenier’s polar factorization theorem (1991)
    - Figalli’s regularity theorem (2010)

语句

在贡献结尾写这一条很不错，(iv) Our experiment results demonstrate the efficiency and efficacy of the proposed method.

最优传输

Brenier interview

- How do you discover the link between optimal transport and hydrodynamics?
- 你对年轻人的建议？
    - To keep some strong personality, avoid to be much follower. Okay, it probably may be a wrong advice. Because it is the most profitable thing to do, to be in a good field, with very good leaders. But i would ... to be free as much as possible. To try to open you our track research.

Brenier lecture: The melting rubik cube: From Fluids to Combinatorics and vice versa.

- 在流体力学中，欧拉是许多著名科学家的跟随者。但是他是第一个，明确地描述流

体的人（1755，用偏微分方程）。
- 不可压缩的流体，被限制在区域D中，并且按照欧拉方程流动。just follows a (constant speed) geodesic curve along the manifold of all possible incompressible maps of D.

一个打乱的魔方，变成液体后流动之后，再还原，变成一个未打乱的模仿。似乎在讲这个意思。

顾险峰老师课件

- Monge问题求最优映射，kantorovich问题是弱化为最优方案.
- 如果最优传输映射存在的话，不要用线性规划去求解。很多0，造成浪费。

Monge问题与Kantorovich问题

Monge问题

$$(MP) \quad \inf\left\{M(T) := \int_X c(x,T(x))d\mu(x) :\ T_{\#}\mu = \nu\right\}$$

note:

$$\nu(A) = \mu(T^{-1}(A))$$

Kantorovich问题连续形式

$$\min_{\gamma} \int_{X\times Y} c(x,y)d\gamma(x,y)$$

相对于γ是一个线性泛函

约束为无限维的凸约束

广义Lagrange乘子法转为对偶问题

广义：约束无限个

ϕ, ψ可以看作Lagrange乘子或者影子价格.

公式演义

惩罚项：

$$
\begin{aligned}
&\forall \gamma \in \mathcal{M}_+(X \times Y) \\
&\qquad \sup_{\phi,\psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X\times Y} (\phi(x)+\psi(y)) d\gamma \\
&\qquad = \begin{cases} 0, \gamma \in \Pi(X,Y) \\ +\infty, \gamma \notin \Pi(X,Y) \end{cases}
\end{aligned}
$$

代入Kantorovich问题

$$$$

Kantorovich原问题（离散形式）

$$
\min_{\gamma} \sum_{i,j}^{m,n} c_{ij}\gamma_{ij}
$$

$$
s.t. \begin{cases} \sum_j \gamma_{ij} \geq \mu_i \\ \sum_i \gamma_{ij} \geq \nu_j \end{cases}
$$

Kantorovich对偶问题（离散形式）

$$
\max_{\phi,\psi} \sum_i^m \phi_i \mu_i + \sum_j^n \psi_j \nu_j
$$

$$
s.t. \begin{cases} \phi_i + \psi_j \leq c_{ij} \\ \phi_i, \psi_j \geq 0 \end{cases}
$$

Weierstrass 定理

f下半连续，X紧，则存在$\bar{x} \in X$,满足

$$
f(\bar{x}) = \min\{f(x) : x \in X\}
$$

Kantorovich问题解的存在性

- 紧空间连续代价函数
- 紧空间下半连续代价函数
- Polish空间（完备的可分度量空间）下半连续代价函数

逐步推广，工程上只用第一种.

Brenier理论 对计算很有用，和微分几何也有联系