CVML

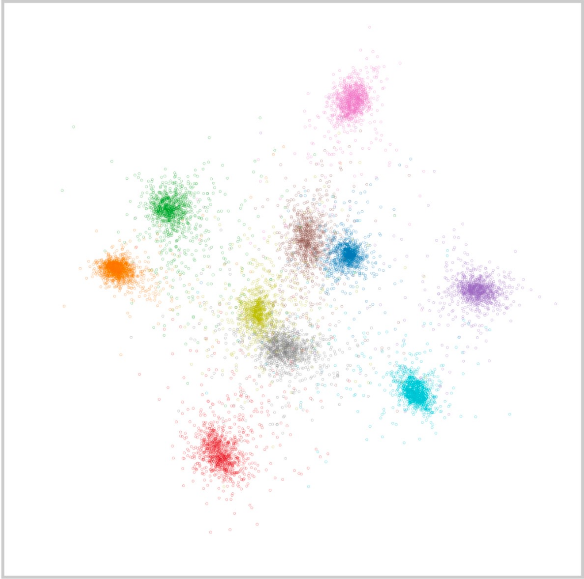
Deep Regression Representation Learning with Topology

Shihao Zhang, Kenji Kawaguchi, Angela Yao

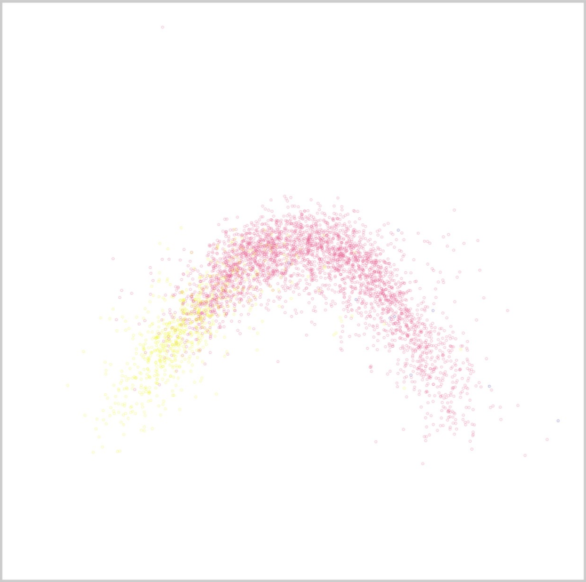
ICML 2024

NUS National University of Singapore | Computing

Motivation



Classification



Regression

The representation topologies of classification and regression are different:

- **Classification**: disconnected
- **Regression**: connected

What topology (shape) the representations should have for effective regression?

Desirable representation

- Intrinsic dimension equals the target space.
- Topologically similar to the target space.

We arrive at this conclusion by establishing two connections:

- $H(Z|Y) \Leftrightarrow$ Intrinsic dimension
- $H(Z|Y),\ H(Y|Z) \Leftrightarrow$ Homeomorphism

Theorem 1: Optimizing the Information Bottleneck $\Longrightarrow$ minimizing $H(Z|Y)$ and $H(Y|Z)$

Information Bottleneck $\Longleftrightarrow$ $H(Z|Y)$, $H(Y|Z)$

Intrinsic dimension equals to the target space

Generalization Error

$$\mathbb{E}_{\{\mathbf{x},\mathbf{z},\mathbf{y}\}\sim P}[||f(\mathbf{z})-\mathbf{y}||_2] \\ \leq \mathbb{E}_{\{\mathbf{x},\mathbf{z},\mathbf{y}\}\sim S}(||f(\mathbf{z})-\mathbf{y}||_2) + 2L_1 Q(\mathcal{H}(\mathbf{Z}|\mathbf{Y}))$$

Generalization error is bounded by $H(Z|Y)$ $\Longrightarrow$ minimizing $H(Z|Y)$ to improve the generalization ability

Intrinsic dimension

$$\mathcal{H}(\mathbf{Z}|\mathbf{Y}) = \mathbb{E}_{\mathbf{y}_i\sim\mathcal{Y}}\mathcal{H}(\mathbf{Z}|\mathbf{Y}=\mathbf{y}_i) \\ \leq \mathbb{E}_{\mathbf{y}_i\sim\mathcal{Y}}[-\log(\epsilon)Dim_{ID}\mathcal{M}_i + \log\frac{K}{C(\epsilon)}]$$

- $H(Z|Y)$ is bounded by the intrinsic dimensions (ID) of M_i $\Longrightarrow$ minimizing the ID of M to lower $H(Z|Y)$
- ID of M should larger than ID of the target space to guarantee sufficient representation capabilities

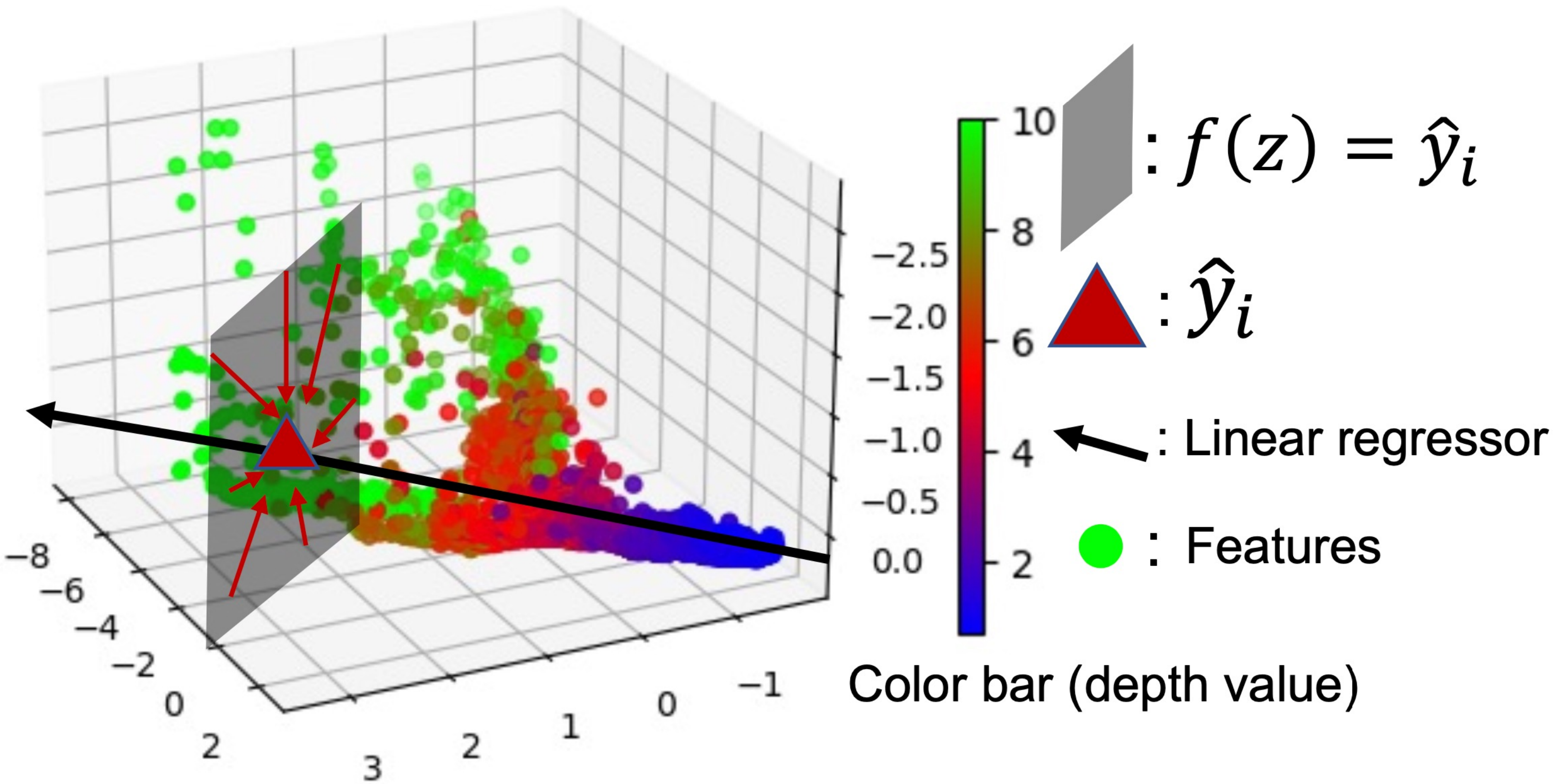
$\Longrightarrow$ ID equals the target space is desirable

Topologically similar to the target space

Definition (Optimal Representation): Z is optimal if $H(Y|Z) = H(X|Z)$ and $H(Z|Y)$ is minimal

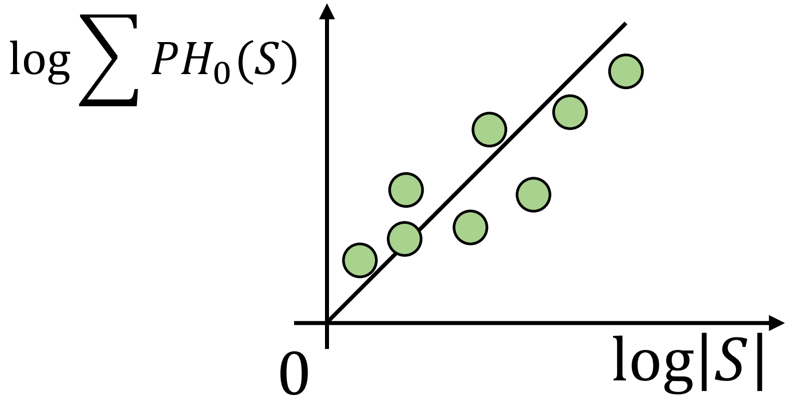
Z is optimal if and only if Z is homeomorphic to Y', where $Y' = Y - N$, N is the aleatoric uncertainty

Encouraging the Same Intrinsic Dimension



Lowering the intrinsic dimension results in a lower $H(Z|\hat{Y})$ *(an approximation for the $H(Z|Y)$),* implying a higher generalization ability.

Encouraging the Same Intrinsic Dimension



Intrinsic dimension can be estimated as the slop between $\log \sum PH_0(S)$ and $\log|S|$[1]

Encourage a lower intrinsic dimension:

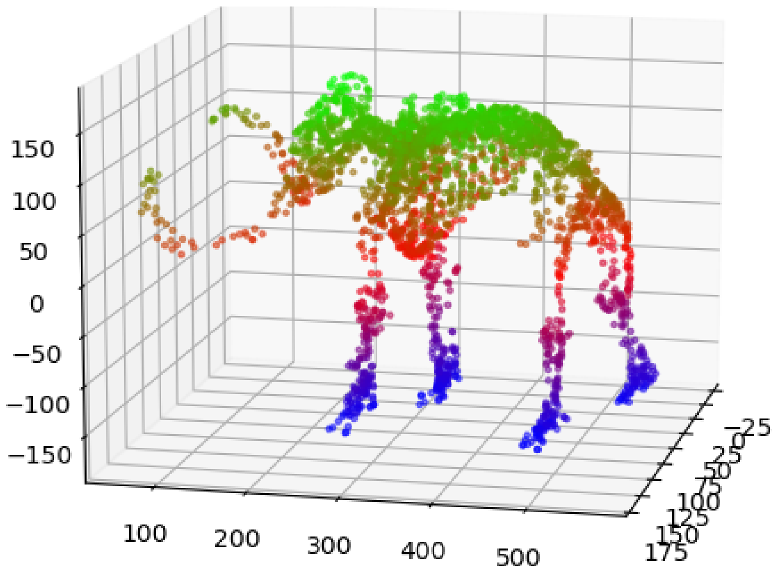
$$L'_d(Z) = slop(\log \sum PH_0(Z), \log|Z|)$$

Encourage the same intrinsic dimension:

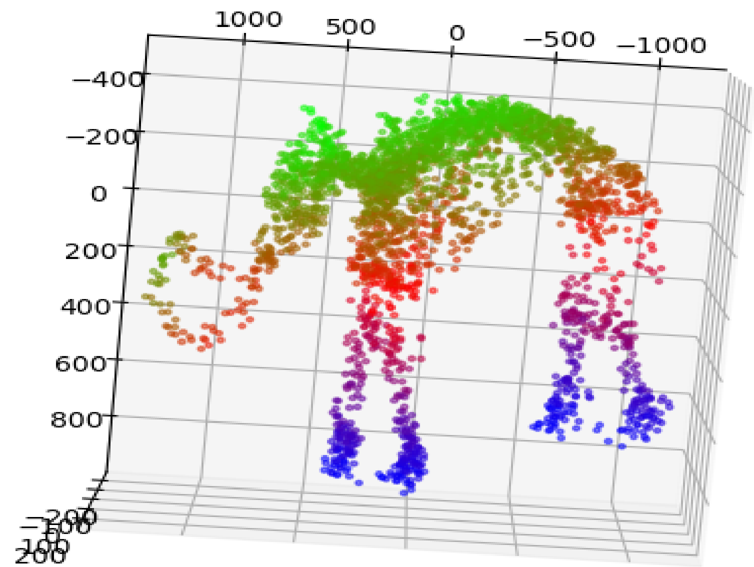
$$L_d = |L'_d(Z)/L'_d(Y)|$$

[1]*Intrinsic Dimension, Persistent Homology and Generalization in Neural Networks, Birdal et al. NeurIPS. 2021*

Enforcing Topological Similarity



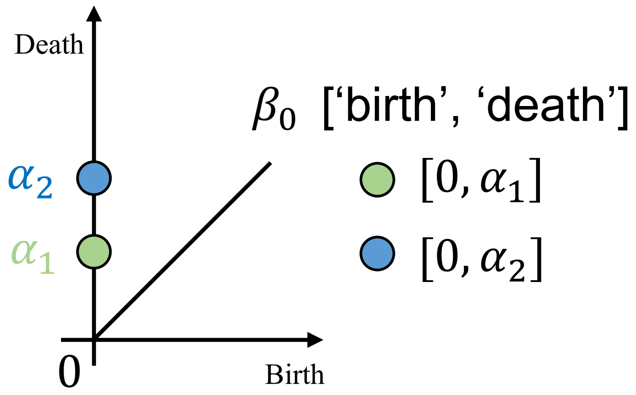
Target Space



Feature Space

Feature and target spaces are topologically similar, and enforcing such similarity is helpful.

Enforcing Topological Similarity



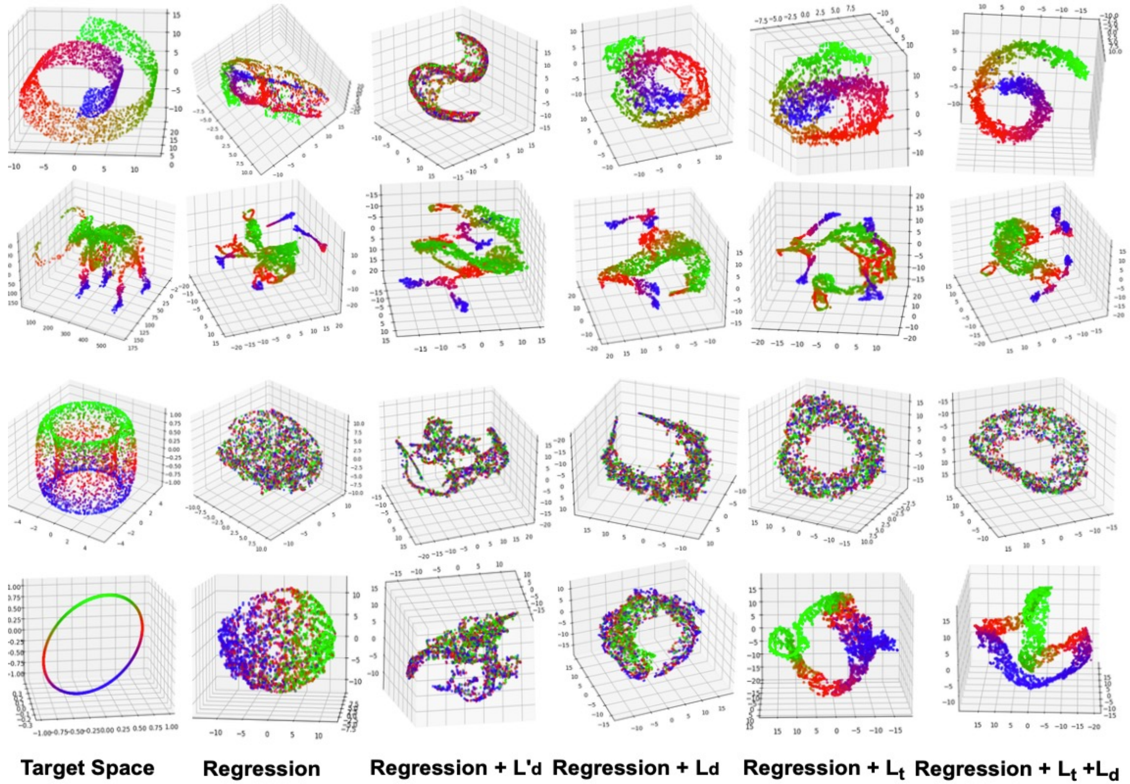
'birth' and 'death' threshold

- The k_{th} persistent homology $PH_k(S)$ is the set of 'birth' and 'death' intervals of the k dimensional holes.
- Calculating $PH_0(S)$ turn out to be find the minimal spanning tree.

Enforcing topological similarity:

$$L_t = ||Z(edge_z) - Y(edge_z)||_2^2 + ||Z(edge_y) - Y(edge_y)||_2^2$$

Topological Autoencoders, Moor et al. ICML. 2021



Method	Swiss Roll	Mammoth	Torus	Circle
Baseline	2.99 ± 0.43	211 ± 55	3.01 ± 0.11	0.154 ± 0.006
+ InfDrop	4.15 ± 0.37	367 ± 50	2.05 ± 0.04	0.093 ± 0.003
+ OE	2.95 ± 0.69	187 ± 88	2.83 ± 0.07	0.114 ± 0.007
$+\mathcal{L}'_d$	2.74 ± 0.85	141 ± 104	1.13 ± 0.06	0.171 ± 0.04
$+\mathcal{L}_d$	0.66 ± 0.08	89 ± 66	0.62 ± 0.12	0.090 ± 0.019
$+\mathcal{L}_t$	1.83 ± 0.70	80 ± 61	0.95 ± 0.05	0.036 ± 0.004
$+\mathcal{L}_d + \mathcal{L}_t$	$\mathbf{0.61 \pm 0.17}$	$\mathbf{49 \pm 27}$	$\mathbf{0.61 \pm 0.05}$	$\mathbf{0.013 \pm 0.008}$

Table 2. Quantitative comparison (MAE) on AgeDB. We report results as mean ± standard variance over 3 runs. **Bold** numbers indicate the best performance.

Method	ALL	Many	Med.	Few
Baseline	7.80 ± 0.12	6.80 ± 0.06	9.11 ± 0.31	13.63 ± 0.43
+ InfDrop	8.04 ± 0.14	7.14 ± 0.20	9.10 ± 0.71	13.61 ± 0.32
+ OE	7.65 ± 0.13	6.72 ± 0.09	8.77 ± 0.49	13.28 ± 0.73
$+\mathcal{L}'_d$	7.75 ± 0.05	6.80 ± 0.11	8.87 ± 0.05	13.61 ± 0.50
$+\mathcal{L}_d$	7.64 ± 0.07	6.82 ± 0.07	8.62 ± 0.20	12.79 ± 0.65
$+\mathcal{L}_t$	7.50 ± 0.04	6.59 ± 0.03	8.75 ± 0.03	12.67 ± 0.24
$+\mathcal{L}_d + \mathcal{L}_t$	**7.32 ± 0.09**	**6.50 ± 0.15**	**8.38 ± 0.11**	**12.18 ± 0.38**

Conclusion

- A desirable representation
  - topologically similar to the target space
  - intrinsic dimension equal to the target space

- Optimizing the Information Bottleneck $\Longrightarrow$ minimizing $H(Z|Y)$ and $H(Y|Z)$
  - $H(Y|Z)$: encourages the representation Z to be informative about the target Y
  - $H(Z|Y)$: can be thought of as noise, and upper-bound the generalization error