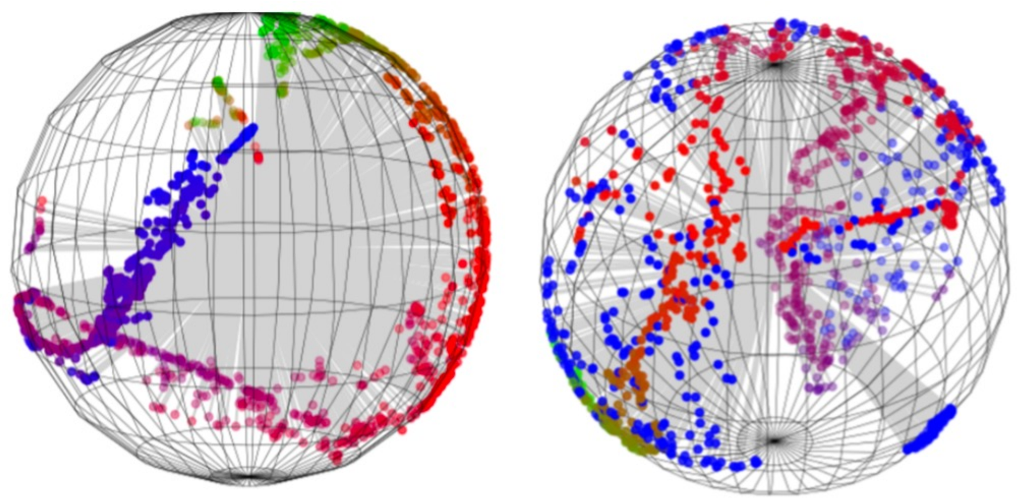
# Deep Regression Representation with Topology

Github page    Project page

Shihao Zhang, Kenji Kawaguchi, Angela Yao
National University of Singapore

NUS | Computing
National University of Singapore

## Motivation



Regression    Classification

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

The representation **topologies** of classification and regression are **different**

Q: Why different topologies?

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

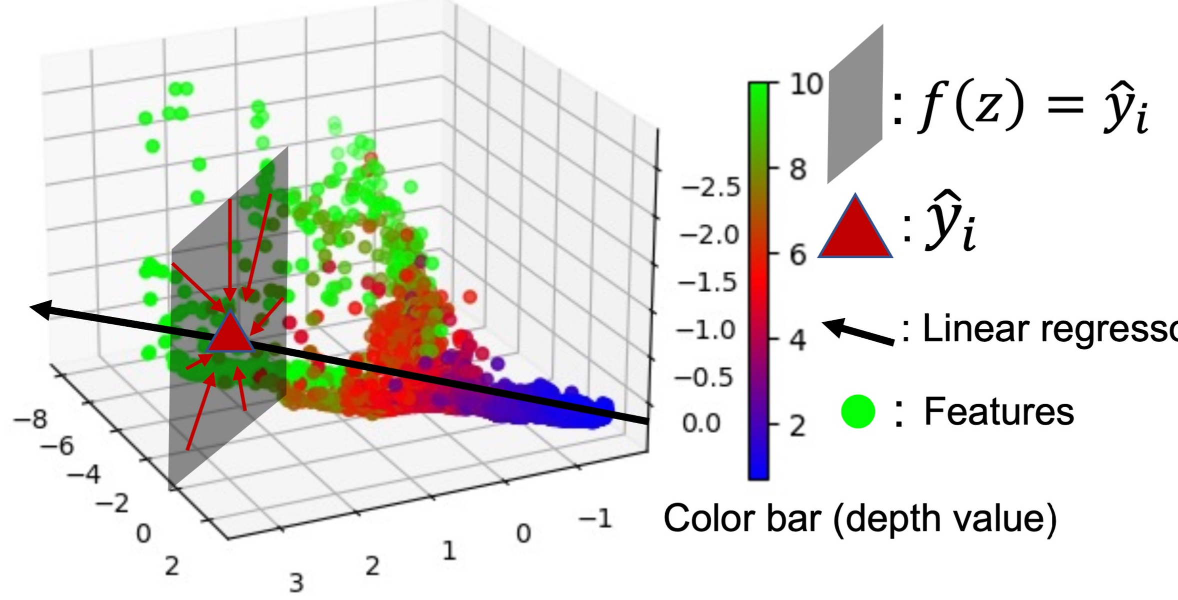
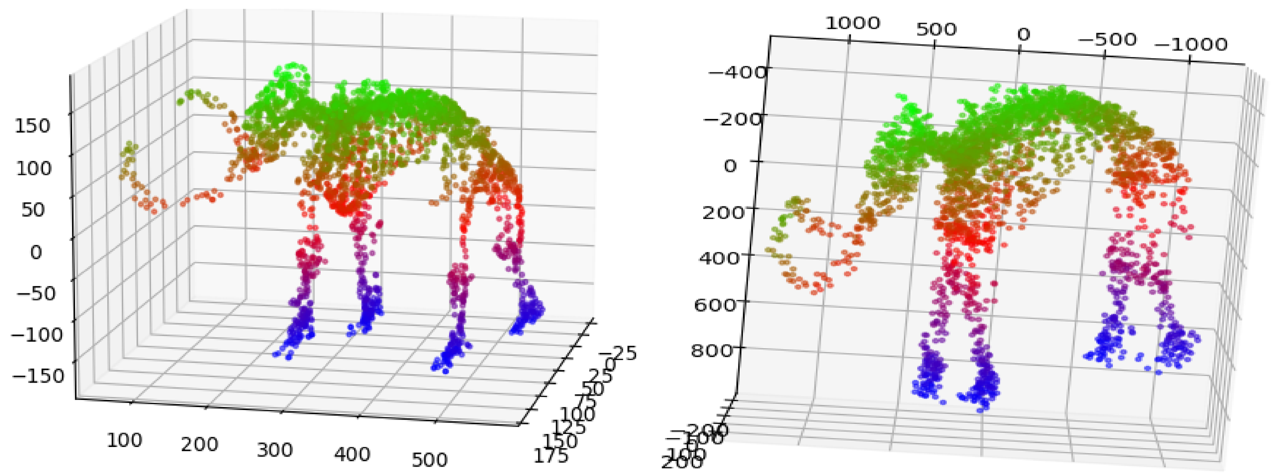



Figure: Visualization of the feature space from depth estimation

Lowering the intrinsic dimension results in a lower H(Z|Y), implying a higher generalization ability



Target Space    Feature Space

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

**Desirable representation**

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

## Theoretical Analysis (informal)

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

Information Bottleneck $\Longleftrightarrow$ $\boxed{H(Z|Y)}$ $H(Y|Z)$

### Intrinsic dimension equals to the target space

**Generalization Error**

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

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

**Intrinsic dimension**

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

- $H(Z|Y)$ 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

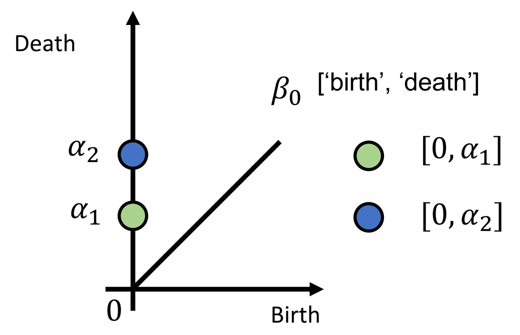
$\Longrightarrow$ ID equals the target space is desirable

### 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

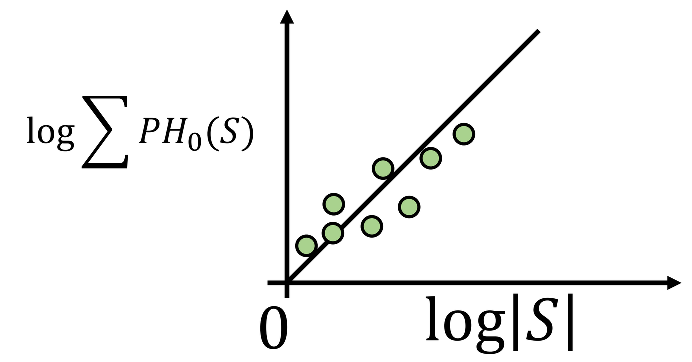
$Z$ is optimal if and only if $Z$ is homeomorphic to $Y'$, where $Y' = Y - N$, $N$ is the aleatoric uncertainty

## Method & Results




'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.
- $edge_s$ : edges of the minimal spanning tree of S
- $PH_0(S)$ can be regarded as the length of the minimal spanning tree of S




Intrinsic dimension can be estimated as the slop between $\log\sum PH_0(S)$ and $\log|S|$ [2]

Encourage a lower intrinsic dimension:

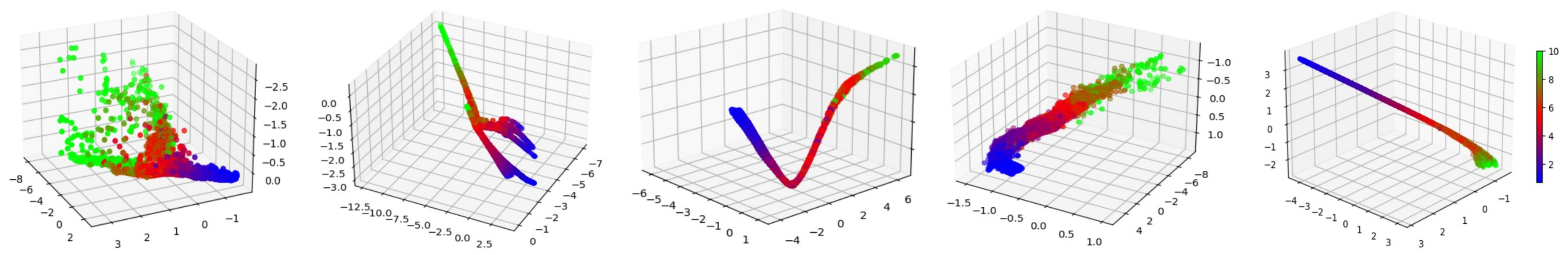
$$L'_d(Z) = slop(\log\sum PH_0(Z), \log|Z|)$$

**Enforcing topological similarity[1]:**

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

**Encourage the same intrinsic dimension:**

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



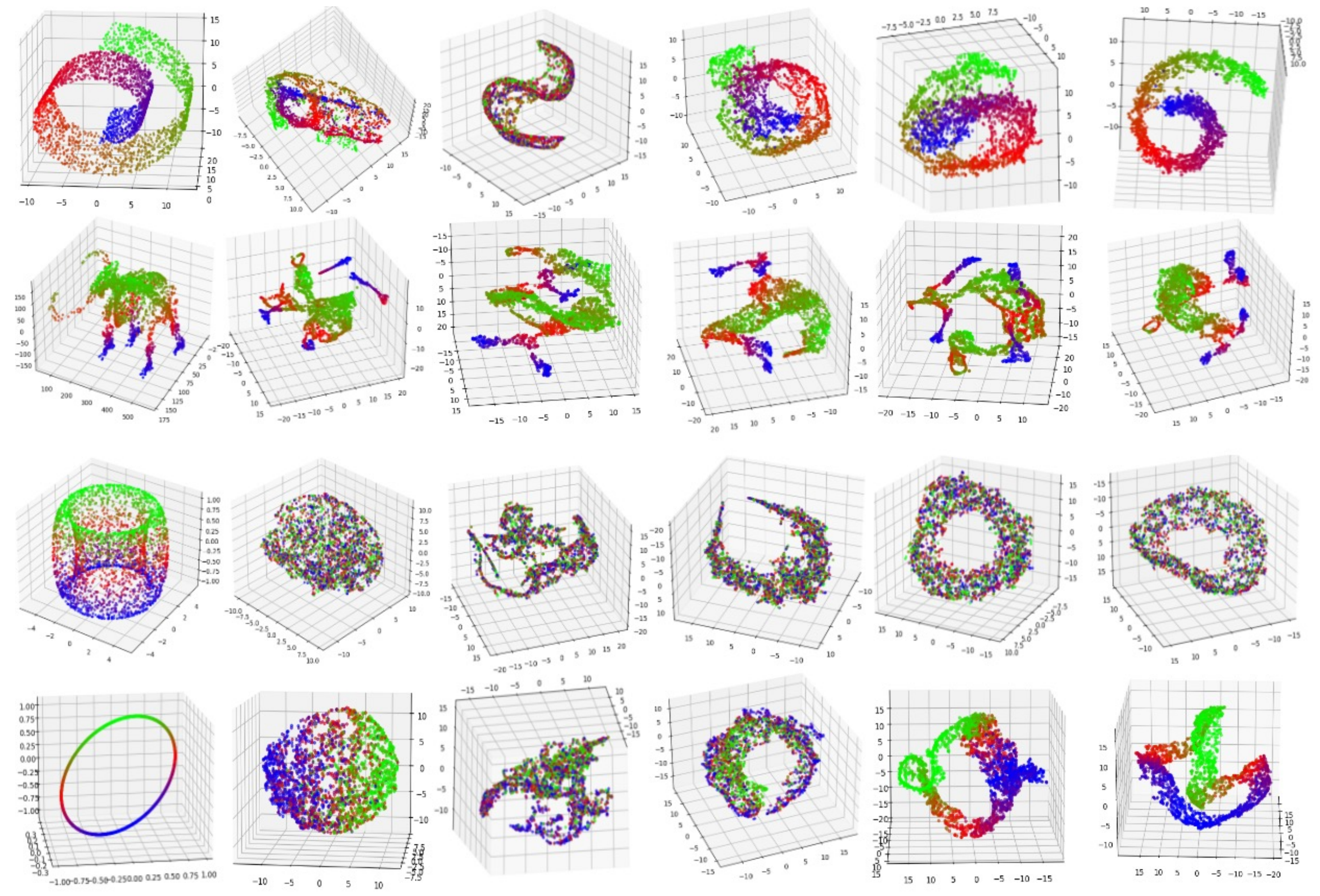
(a) Regression    (b) Regression $+\mathcal{L}'_d$    (c) Regression $+\mathcal{L}_d$    (d) Regression $+\mathcal{L}_t$    (e) Regression $+\mathcal{L}_R$

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

*Table 5.* Quantitative comparison of the time consumption and memory usage on the synthetic dataset and NYU-Depth-v2, and the corresponding training times are 10000 and 1 epoch, respectively.

| $n_m$ | Regularizer | Coordinate Prediction (2 Layer MLP) | | Depth Estimation (ResNet-50) | |
|---|---|---|---|---|---|
| | | Training(s) | Memory (MB) | Training(s) | Memory (MB) |
| 0 | - | 8.88 | 959 | 1929 | 11821 |
| 100 | $\mathcal{L}_t$ | 175.06 | 959 | 1942 | 11833 |
| 100 | $\mathcal{L}_d$ | 439.68 | 973 | 1950 | 12211 |
| 100 | $\mathcal{L}_t + \mathcal{L}_d$ | 617.41 | 973 | 1980 | 12211 |
| 300 | $\mathcal{L}_t + \mathcal{L}_d$ | - | - | 2370 | 12211 |



Target Space    Regression    Regression + L'd    Regression + Ld    Regression + $L_t$    Regression + $L_t$ + $L_d$

*Table 1.* Results ($\mathcal{L}_{\text{mse}}$) on the synthetic dataset. We report results as mean ± standard variance over 10 runs. **Bold** numbers indicate the best performance.

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

References

[1] Moor et al. Topological Autoencoders. ICML. 2021
[2] Birdal et al. Intrinsic Dimension, Persistent Homology and Generalization in Neural Networks. NeurIPS. 2021