Improving Deep Regression with Ordinal Entropy

Shihao Zhang, Linlin Yang, Michael Bi Mi, Xiaoxu Zheng, Angela Yao

April 9, 2023

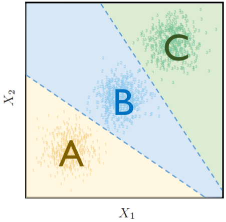
Machine Learning (ML)

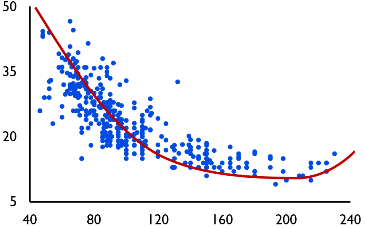
Classification and regression are two fundamental tasks of ML.

Classification vs. Regression

Categorical target

Continuous target





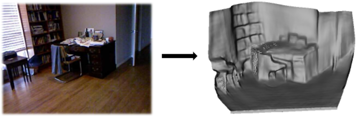
Figures adapted from Elements of Statistical Learning 2^{nd} Ed. by Hastie et al.

Classification in Computer Vision Performs Better

Age Estimation

Counting

Depth Estimation



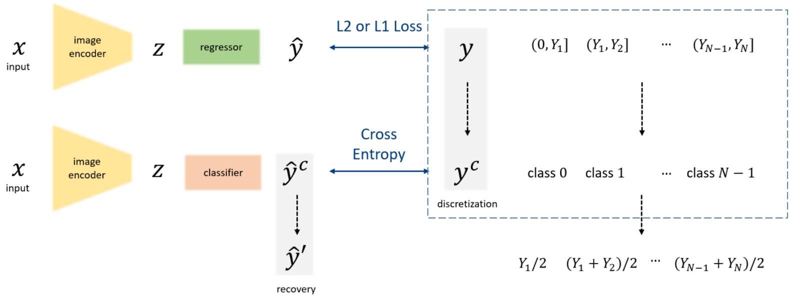
Others

Pose estimation
Optical flow
Super-resolution
...

Formulating Regression as a Classification Task

Disadvantages of formulating regression as a classification task:

- Introduce discretization errors.
- Lose ordinal relationship.



Observation

Observation

Despite the disadvantages, **classification outperforms regression**, for a diverse set of regression problems.

Why?

Why Classification Outperforms Regression

Task Specific Reasons:

- **Depth estimation**: predicting depth range is easier
- **Pose estimation**: bias + problematic gradients
- **Local counting**: ambiguous input & ground truth labels

Research Problem

Anything more general that is not task specific?

Feature Space of Classification

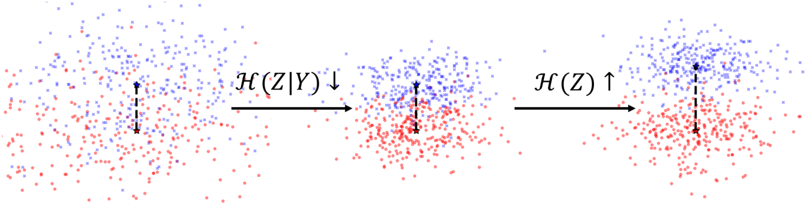
$$\underbrace{\mathcal{I}(Z;Y)}_{\textit{Mutual Information}\uparrow} = \underbrace{\mathcal{H}(Z)}_{\textit{Marginal Entropy}\uparrow} - \underbrace{\mathcal{H}(Z|Y)}_{\textit{Conditional Entropy}\downarrow}$$

Intra-class feature distances $\downarrow$:

- $\Rightarrow$ lower-entropy clusters, $\mathcal{H}(Z|Y)\downarrow$.

Inter-class feature distances $\uparrow$:

- $\Rightarrow$ higher-entropy feature space, $\mathcal{H}(Z)\uparrow$.

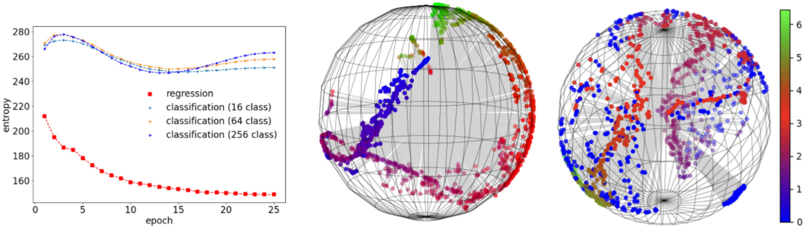


Motivation

Yet, regression has no concept about class

- **What the feature space of regression looks like?**

Feature Space: Regression VS. Classification



(a) Entropy of feature space (b) Regression (c) Classification

We show that MSE loss is a proxy only for minimizing $\mathcal{H}(Z|Y)$.

- $\Rightarrow$ Regression with MSE lags in its ability to maximize $\mathcal{I}(Z;Y)$ by learning high-entropy feature space.

Regularizers to Increase Feature Entropy

Encourage a higher-entropy feature space in regression by minimizing the negative distances between feature centers $\mathbf{z}_{c_i}$:

$$\mathcal{L}'_d = -\frac{1}{M(M-1)} \underbrace{\sum_{i=1}^{M} \sum_{i \neq j} ||\mathbf{z}_{c_i} - \mathbf{z}_{c_j}||_2}_{\propto \mathcal{H}(Z)}, \tag{1}$$

Regularizers to Increase Feature Entropy

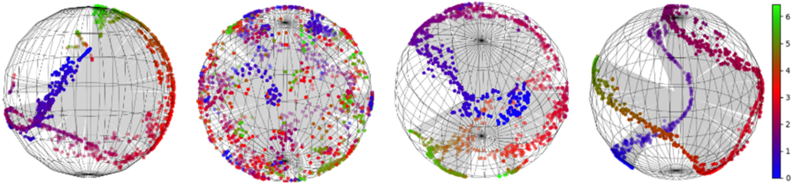
$\mathcal{L}'_d$ may break the ordinality in the feature space. To preserve ordinality:

$$\mathcal{L}_d = -\frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{i \neq j} w_{ij} ||\mathbf{z}_{c_i} - \mathbf{z}_{c_j}||_2, \qquad \text{where } w_{ij} = ||y_i - y_j||_2 \tag{2}$$

To further minimize the conditional entropy $\mathcal{H}(Z|Y)$:

$$\mathcal{L}_t = \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{z}_i - \mathbf{z}_{c_i}||_2. \tag{3}$$

Tradeoff in Entropy vs. Ordinality



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

- High entropy feature space is good for a higher $\mathcal{I}(Z, Y)$.
- Increasing entropy naively destroys ordinal relationships.
- Preserving ordinality again lowers entropy.

Conclusion

- We show that regression with MSE lags in its ability to maximize $\mathcal{I}(Z, Y)$ by learning high-entropy features space.
- We design an ordinal entropy regularizer to learn high-entropy feature representations which preserve ordinality for regression.

Future Works

- Why not solve everything as classification?

Thanks