Box-valued AE - a single framework for denoising, (self/semi/un)supervised and transfer learning

Box as an uncertainty model and AE input/output

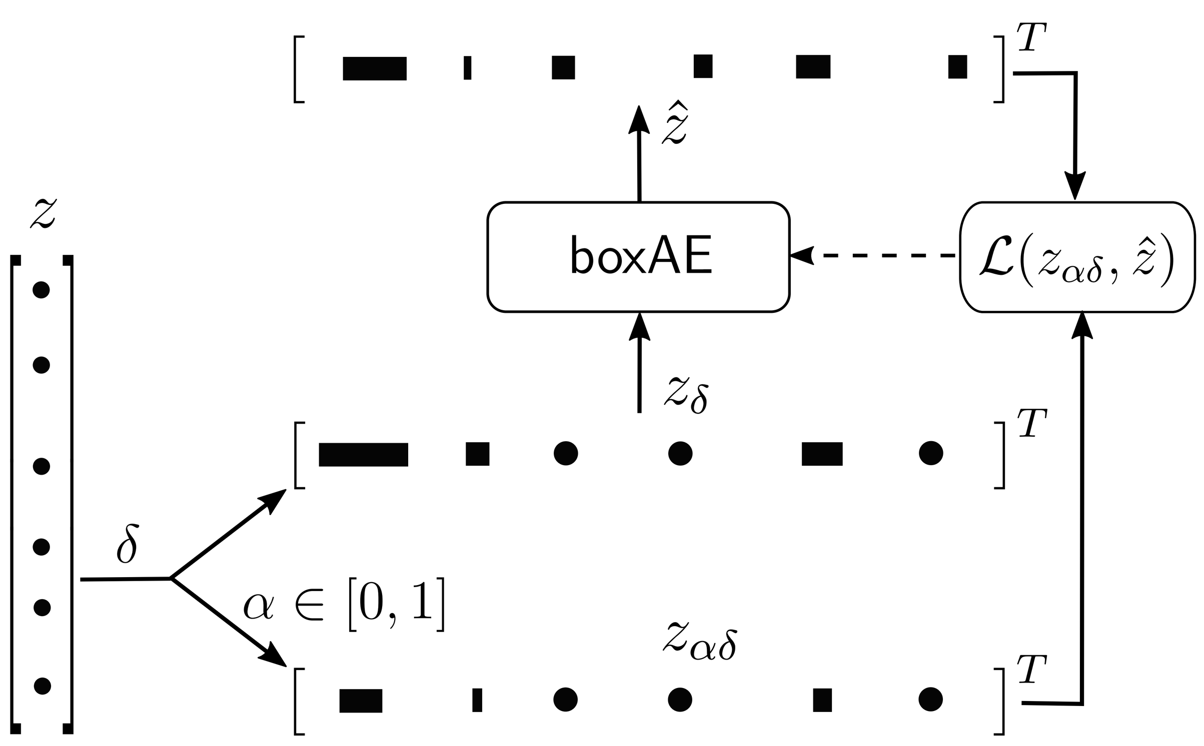
Mladen Đalto, mladendalto@gmail.com

Motivation

- reduce discarded data
- enable use of partially matching datasets

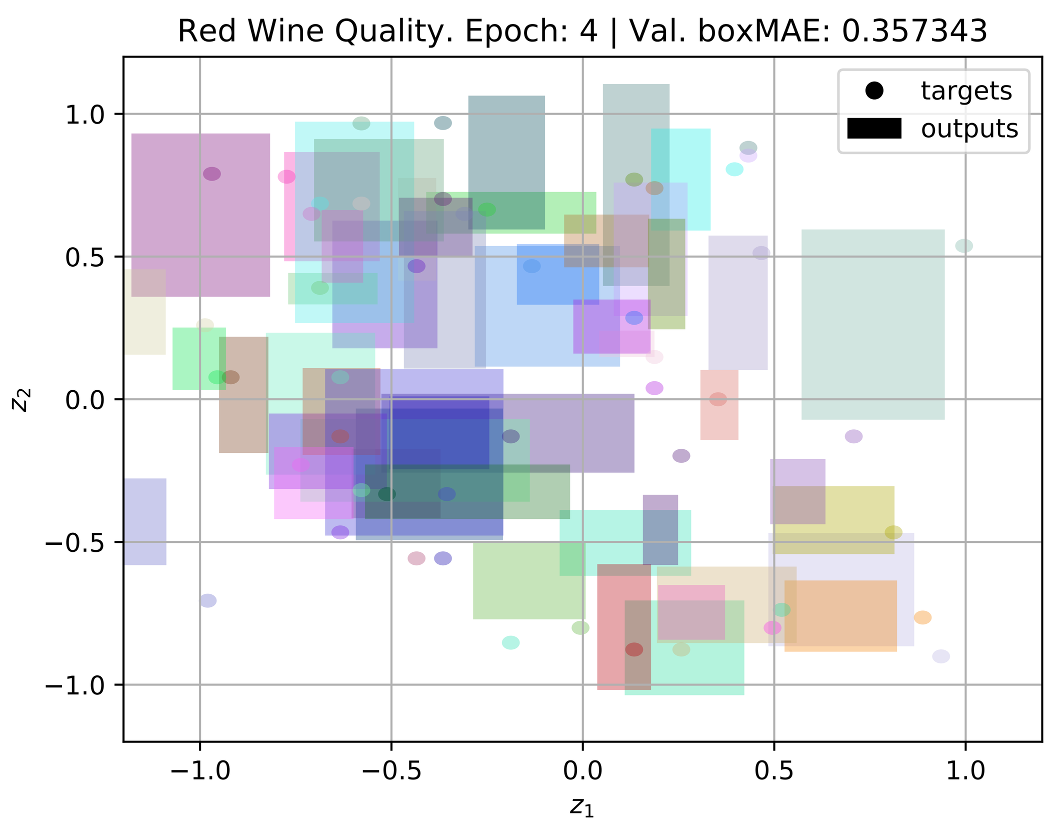
Theory

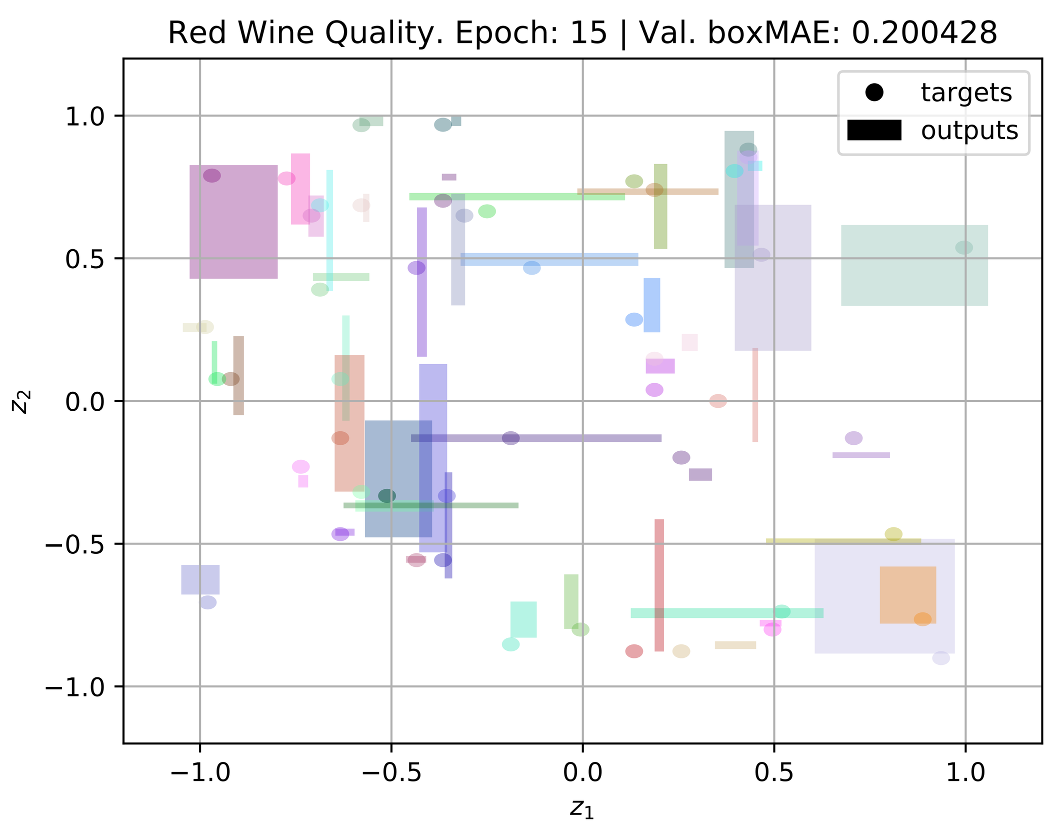
- conditions ensuring: box $\rightarrow$ AE $\rightarrow$ box, (margin)
- uncertain or missing samples represented as subset of e.g. $[-1,1]^n$
- morphing corruption, $\delta \sim \mathbb{U}_{[0,1]^n}$: $z \rightarrow z_\delta = (1_n - \delta) \odot z \oplus \delta\,[-1,1]^n$
- independent corruption, $\delta_u, \delta_d \sim \mathbb{U}_{[0,1]^n}$: $z \rightarrow z_{\delta_d,\delta_u} = [(1_n - \delta_d) \odot z + \delta_d \odot d,\ (1_n - \delta_u) \odot z + \delta_u \odot u]$
- loss = any box distance + λ box volume: $\mathcal{L}_p(\mathcal{Z}_k, \hat{\mathcal{Z}}_k) = \max(0, \hat{u}_k - u_k)^p + \max(0, d_k - \hat{u}_k)^p + \max(0, \hat{d}_k - u_k)^p + \max(0, d_k - \hat{d}_k)^p + \lambda|\hat{u}_k - \hat{d}_k|^q$



Results

- conditions for box $\rightarrow$AE$\rightarrow$ smaller box, (margin)
- work in progress with examples (classification $\rightarrow$ theory only)





corrupted samples	$\max\delta$/ dataset	0.5	0.4	0.3	0.2	0.1	0
1	CCPP	0.1492	0.1191	0.0929	0.0640	0.0349	0.004
	RWQ	0.0944	0.0749	0.0571	0.0475	0.0375	0.0172
	EE	0.1699	0.1588	0.0804	0.0814	0.1320	0.0484
2	CCPP	0.2419	0.1938	0.1526	0.1026	0.0544	0.004
	RWQ	0.1430	0.1117	0.0857	0.0618	0.0381	0.0172
	EE	0.2279	0.1856	0.1189	0.1046	0.1150	0.0484
3	CCPP	0.3269	0.26847	0.2053	0.1417	0.0724	0.004
	RWQ	0.1854	0.1494	0.1136	0.0768	0.0441	0.0172
	EE	0.2648	0.1848	0.1564	0.1027	0.0954	0.0484
4	CCPP	0.4095	0.3354	0.2594	0.1787	0.0914	0.004
	RWQ	0.2219	0.1768	0.1373	0.0928	0.0493	0.0172
	EE	0.3361	0.2597	0.2078	0.147	0.0947	0.0484

Conditions ensuring box $\rightarrow$ AE $\rightarrow$ box

Matrix-box multiplication can be written simpler if we decompose $W \in \mathbb{R}^{m\times n}$ into a difference of matrices with non-negative elements $W = W^+ - W^-$ where , $W^+ = \max(W, 0)$, and $W^- = \max(-W, 0)$, giving a box $W\mathcal{R} =$ $\{z \in \mathbb{R}^m \,|\, W^+ d - W^- u \preceq z \preceq W^+ u - W^- d\}$.

Another operation which preserves boxes is translation by a vector $b \in \mathbb{R}^m$:

$$\mathcal{R} \oplus b = \{z \in \mathbb{R}^m \,|\, d + b \preceq z \preceq u + b\}$$

where the Minkowski sum is taken between a box $\mathcal{R}$ and a singleton set b.

If we assume monotonic activation function and apply it to a box we get:

$$\psi(\mathcal{R}) = \begin{cases} \psi(d) \preceq z \preceq \psi(u), \text{if } \psi' \geq 0, \\ \psi(u) \preceq z \preceq \psi(d), \text{if } \psi' \leq 0. \end{cases}$$

box $\rightarrow$ AE $\rightarrow$ smaller box?

Box-valued AE with L hidden layers can be formulated as follows:

$$\begin{aligned} \mathcal{Z}_0 &= \mathcal{X}_k \times \mathcal{Y}_k, \\ \mathcal{V}_l &= W_l \mathcal{Z}_{l-1} \oplus \{b_l\}, 1 \leq l \leq L \\ \mathcal{Z}_l &= \psi_l(\mathcal{V}_l), 1 \leq l \leq L \\ \hat{\mathcal{Z}} &= \mathcal{Z}_L, \end{aligned}$$

where $\psi_l(\cdot)$ are monotonic.

Volume of the boxes propagated through boxAE layers is a measure of uncertainty. Generally, we have no analytic expression for $|\hat{\mathcal{Z}}|$ in terms of W_l so we resort to upper bounds derived using Lipschitz continuity:

$$|\hat{\mathcal{Z}}| \leq \left(\prod_{l=1}^{L} K_l^{n_l} \sqrt{|\det(W_l^T W_l)|}\right) |\mathcal{X}_k||\mathcal{Y}_k|$$

where K_l is the Lipschitz constant of $\psi_l(\cdot)$. **When $|\hat{\mathcal{Z}}|/(|\mathcal{X}_k||\mathcal{Y}_k|) \leq 1$ boxAE will reduce box volume almost everywhere.**

Datasets:
CCPP - Combined Cycle Power Plant
RWQ - Red Wine Quality
EE - Energy Efficiency

Visage Technologies

Download the poster $\rightarrow$