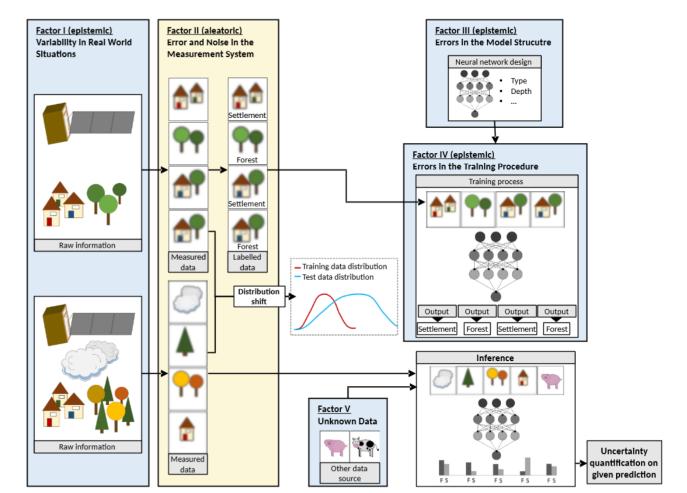
A MULTIOBJECTIVE VIEW ON CREATING COUNTERFACTUAL EXPLANATIONS FOR EXPLAINING UNCERTAINTY IN MACHINE LEARNING

EUROPT 2022, 30th July 2022 Julia Niebling German Aerospace Centre, Institute of Data Science, Jena, Germany



Uncertainty in Machine Learning

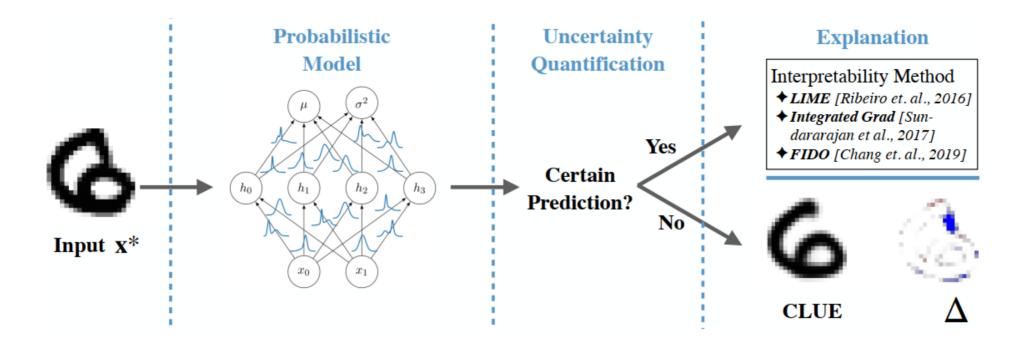
- Quantification of uncertainty to
 - Understand models
 - Trust predictions
 - Develop robust methods
 - Aleatoric (data) uncertainty Epistemic (model) uncertainty



Gawlikowski et.al. A Survey of Uncertainty in Deep Neural Networks

Getting a CLUE





CLUE = Counterfactual Latent Uncertainty Explanations

Counterfactual Explanation = what should be different to change the outcome

Antorán et.al. Getting a CLUE: A Method for Explaining Uncertainty Estimates (ICLR 2021)

Paraphrasing Optimization \Leftrightarrow Machine Learning



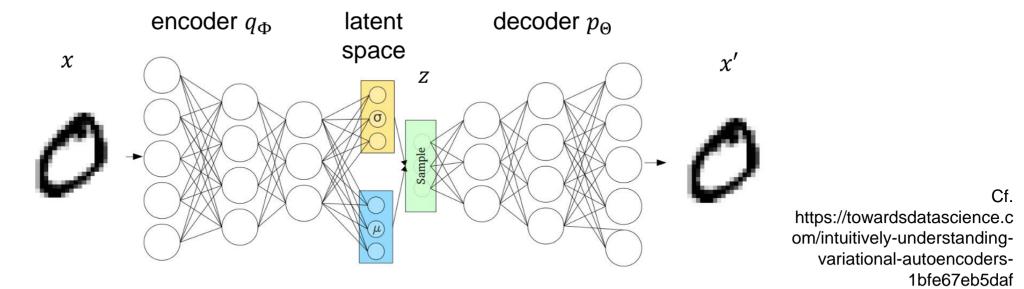
- Optimization is used in ML to train models, e.g. neural networks
- Further usage of optimization in ML possible

OPTIMIZATION	MACHINE LEARNING
objective function	loss function
to optimize	to train
iteration (e.g. of gradient descent)	epoch
$x \in \mathbb{R}^n$ (variables in search space)	e.g. $\theta, w \in \mathbb{R}^n$ (parameters of a NN when <i>training a NN</i>)

Notations



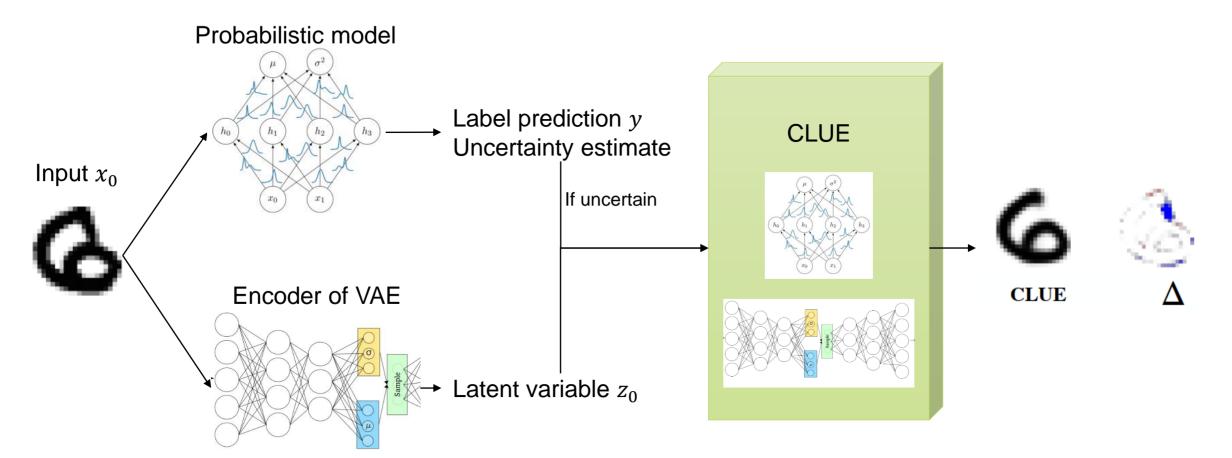
- data points x, labels y, latent space variables z
- probabilistic model \mathcal{H} , differentiable, to estimate uncertainty of an observation, e.g. Bayesian NN
- variational autoencoder (VAE) consisting of encoder $q_{\Phi}(x|z)$ and decoder $p_{\Theta}(x|z)$ with parameters/weights Φ, Θ



• predictive means of decoder and encoder of VAE: $E_{p_{\Theta}(x|z)}[x] =: \mu_{\Theta}(x|z), E_{q_{\Phi}(z|x)}[z] =: \mu_{\Phi}(z|x)$

Pipeline





6

The optimization problem in CLUE

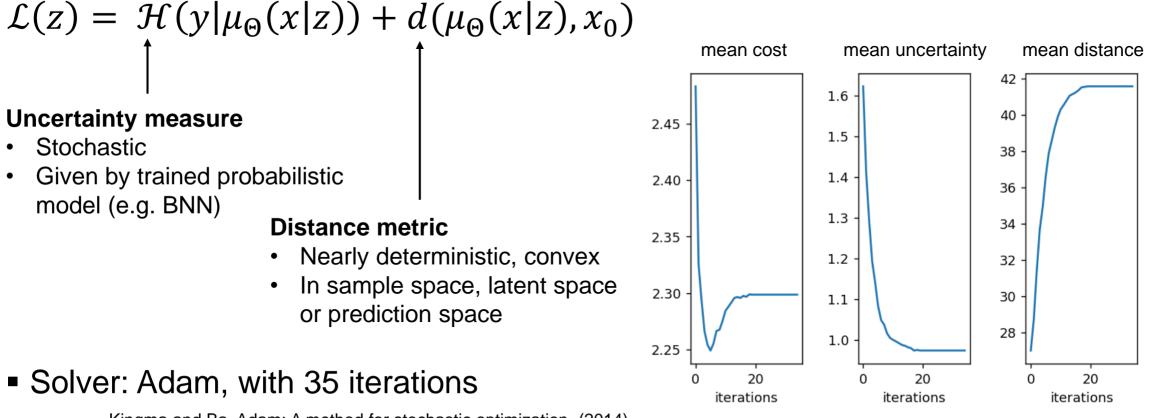


- Stochastic
- Given by trained probabilistic model (e.g. BNN)

Distance metric

- Nearly deterministic, convex
- In sample space, latent space ۲ or prediction space
- Solver: Adam, with 35 iterations

Kingma and Ba. Adam: A method for stochastic optimization. (2014)





The multiobjective optimization problem in CLUE

 $\mathcal{L}(z) = \mathcal{H}(y|\mu_{\Theta}(x|z)) + d(\mu_{\Theta}(x|z), x_0) \quad \clubsuit$

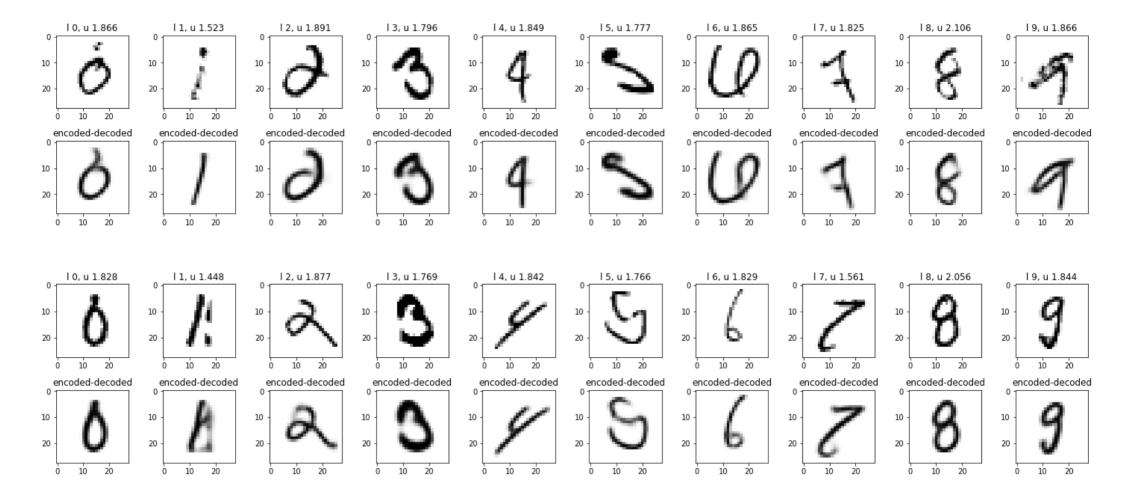
- Two objectives summed up with weight 1
- Different weights to
 - Indicate preferences of decision maker
 - Give more insights to possible solutions
 - ...?
- Solving with weighted sum approach

$$\begin{split} \mathcal{L}(z) &= \lambda \cdot \mathcal{H}(y | \mu_{\Theta}(x | z)) + (1 - \lambda) \cdot d(\mu_{\Theta}(x | z), x_0), \\ \lambda \in \{0, 0.1, 0.2, \dots, 1\} \end{split}$$

$$\mathcal{L}(z) = \begin{pmatrix} \mathcal{H}(y|\mu_{\Theta}(x|z)) \\ d(\mu_{\Theta}(x|z), x_0) \end{pmatrix}$$

Considered uncertain test samples



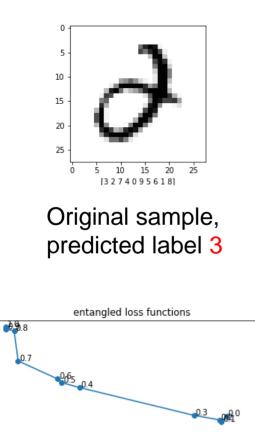


Exemplary results for number 2

1.75

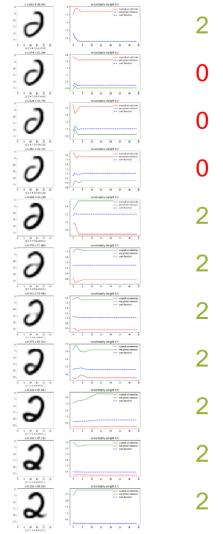
1.50

ィ



iterations
$= \int_{a}^{a} \int_{a}^{b} \int_{a}^{b} \int_{a}^{a} \int_{a}^{b} \int_$
$= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\$
$= \underbrace{\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$
$= \underbrace{\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$
$= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\$
$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1$
$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1$





. . .

80

- 09 - 05 - 05

20

0

0.25

0.50

0.75

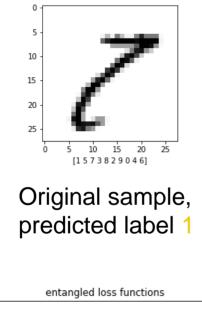
1.00

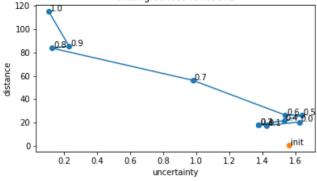
uncertainty

1.25

Exemplary results for number 7

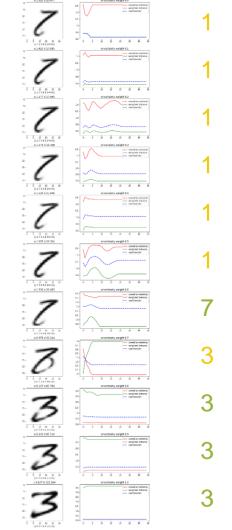
イ





3 3 3 3 3 3 3 3 3 3





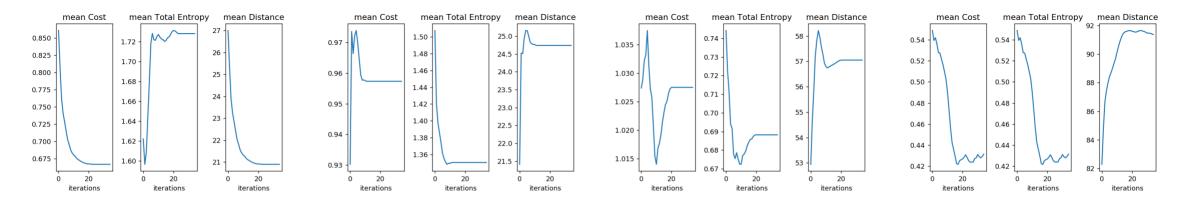
. . .

Loss functions and interpretation



 $\lambda = 1 0$





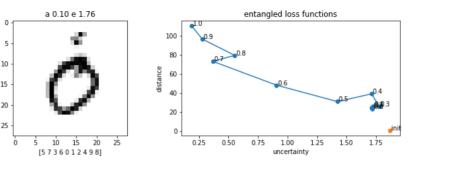
 $\lambda = 0.7$

- More weight on distance: more similar CLUEs, often no change in predicted label
- More weight on uncertainty: less similar CLUEs, change in label possible
- Iocal solutions!
- first iteration (1 step of Adam) step brings most visually
- too much weight to one objective often not that useful



- More extensive studies, different data sets
- Solving methods for multiobjective optimization problems
 - Local:
 - Better initialization/starting points for CLUE → make use of previous computed solutions
 - Stochastic gradient descent methods for multiobjective problems (?)
 - Global: Evolutionary algorithms, e.g. NSGA-II

Thank you for your attention!



DLR

Further examples

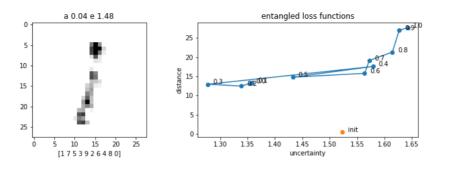
10

15

20

25

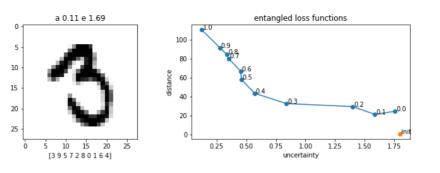
б б б б б б б Ь б б б Ь Ь д Ь б б Ь б Ь Ь Ь Ь б б б б б Ь ò ò б Ь Ó Ó Ò Ò Ò Ò Ò Ò Ò Ò Ò Ò Ò Ò Ò Ò Ò Ò Ó Ó Ó Ô Ò Ò Ô Ô Ó 0 0 Ô 0 Ô Ô 0 Ô Ô 0 Ó Ò Ò Ò Ó Ò Ô 0 Ó Ò 0 0 0 0 0 0 0 0 0 0 0 0 Ô 0 0 0 0 0 0 0 0 0 0 0 0 Ò 0 Ò Ó Ó Ó Ó Ò Ò Ò Ò Ò Ó Ò Ò Ò Ó Ò Ò Ó Ò Ó Ó Ò Ò Ò Ò Ó Ò Ò Ò 0 0 0 Ô Ò Ò Ò Ò Ъ Ь Ь Ò Ò Ь 0 0 Ò Ó Ò Ò Ó 0 0 0 Ó 0 Ó 0 0 0 Ô 0 Ó 0 Ó Ó 0 0 Ó Ó Ó Ó 0 0 Ó Ó Ó Ó 0 0 0 0 Ó Ó 0 0 0 0 0 Ó 0 0 0 0 0 Ó Ó Ó Ó Ó 0 0 \circ 0 Ø 0 0 0 Ø 0 0 6 0 0 b Ь 0 6 b 0 Ø Ø Ó Ø 0 Ь Ø Ø Ø Ø Ø ß D D D D Ø 0 6 B B B B B B В B B В B ß В В В в B В В ß Ь Ь в ĸ В В В ĸ ß в ß ß ß B 0 O B Ø O ß B B 0 Ø B B Ø Ø B Ø ß B B Ø Ø B B B Ø B B B ß Ø Ø Ø Ø Ď O 0 0 0 Ø 0 Ø 0 Ø 0 00 B 0 0 0 Ø Ø 0.0.0. 0 Ø 0 Ø O Ø 0 0 O 0 0 O O O



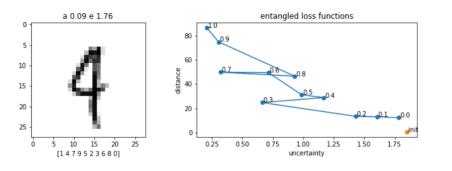


30 -33 -10 30 -< |¹/₂ 1.334. » 1 » 1 10 1 20 1 1 . 1 . 1 : 1 10 20 -10 33 10 20 20 10 33 20 -33 -10 -20 -: 1 20-10 -32 -20 -10 10 20 20 · 13 · 20 · 10. 20-20. 10 -20 20 -10. 10 -20. 30 10. 20 -30 -: / e 10 20 30 32 -10 30 10 20 -10 -20 -10. 33 -20 -20-30 20 -30 30 -20 -32 -10 -20 -20 10 · 30 10 20 -10 1) 1) 2) [10-20-20 30 - 10 - 20 - 20 -[10 - 10 20 - 20 -[I I [20 -20 -20 -20 -20 -33 -20 -20 -30 · 28 · 30 -20 -1 10 20-20-/ 10-20-20-1 1 20. 10. 10. 10. 10. 11. 10. 10. 10. 10. 11. 10. 10. 10. 10. 10. 10. Ţ 20-20. 2. 20. 10. 20. 20. 1 2.) ». 2. 2. 2. Ţ 31-20 · 20 · 28 · 28 · ſ 1 ... Ţ 30 -20 -10 1 1 1 - 1 1 1 1 ... 1 1 1 2 1 2 1 20. 1 20. 1 . 1 1 1 . 1 1 . 1 . 1 10 . 1 1 1 ... 213 - 113 - 0 20-20-20-20-20-20-1 ». J .] 11-12-20-20-20-20-20-1 20 - 10 -20 - 20 -5 5



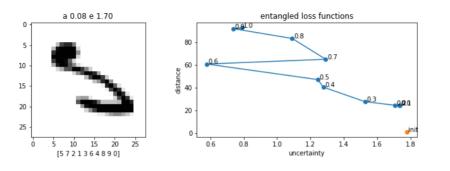


	$\begin{array}{c} 1.566 \ d 21.600 \\ \hline \\ \hline \\ \hline \\ \hline \\ \\ \hline \\ \\ \\ \\ \\ \\ \\ \\$
	$\begin{array}{c} 0.270 \ 4.2 \ 899 \\ \hline \begin{array}{c} 0.270 \ 4.2 \ 899 \\ \hline \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \hline \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} 0 \\ \end{array} \\$
	$ \begin{array}{c} 0.450 d 50 531 \\ \hline 5 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10$
VOLUMENT VALUE VAL	$ \begin{array}{c} & 0.330 \text{ d } 48353 \\ & 5 \\ & 5 \\ & 5 \\ & 5 \\ & 5 \\ & 6 \\ & 5 \\ & 6 \\ & 5 \\ & 6 \\ & 5 \\ & 6 \\ & 6 \\ & 7 $
$ \frac{1}{2} 1$	$ \begin{array}{c} & 0 \\ $
	$\underbrace{\sum_{j=1}^{n-1} \frac{1}{j}}_{j=1}^{n} \sum_{j=1}^{n-1} \frac{1}{j} \underbrace{\sum_{j=1}^{n-1} \frac{1}{j}}_{j=1}^{n-1} \frac{1}{j} \underbrace{\sum_{j=1}^{n-1} \frac{1}{j}}_{j=1}^{n-1} \underbrace{\sum_{j=1}^{n-1} \frac{1}{j}}}_{j=1}^{n-1} \underbrace{\sum_{j=1}^{n-1} \frac{1}{j}}}_{j=1}$



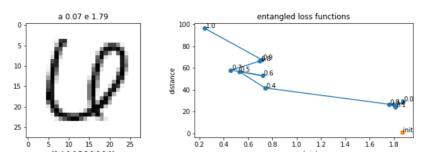


n 1.785 d 12.295 1 410 4 13 40 u 0.661 d 24.77 Δ ム ム Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Л Δ Λ Δ Л Л Л Л Л Л Л Л Л Л Л Л Δ Δ Δ Λ



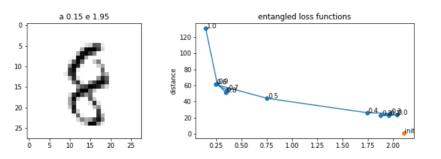


u 1 527 d 27 588 * ¢, ¢ ¢, < S رک :5 » **S** < ٩, <



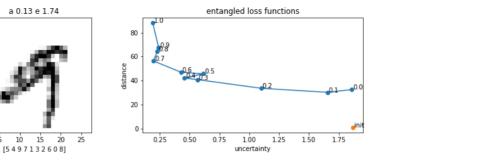


0 :0 0 0 0 0 0 (0 0 0 0 0 0 0 0 0 0 10 0 (0 0 0 (0 10 0 0 0 0 10 0 0 0 0 :0 U 0 0 0 0 0 Ø 10 0 0 0 (0 10 Ø Ø (0 0 0 0 0 0 0 U 0 0 0 Ø 0 0 10 0 0 U Ø 0 U 0 0 U Ø 10 0 (0 0 Ø U 0 0 10 0 U U 0 U U 0 U 0 0 0 U 0 U U :0 U U U U U U 0 0 U 0 U U U U U U U U U U 0 \mathcal{O} 0 \mathcal{O} 0000 \mathcal{O} \mathcal{O} 0000 00 0 \mathcal{O} 0 0 00 0 0 0 0 0 O U \mathcal{O} O 0 0 0 0 $\overline{\mathcal{O}}$ \mathcal{O} \mathcal{O} \mathcal{O} \mathcal{O} \mathcal{O} \mathcal{O} 00 U U \mathcal{O} \mathcal{O} \mathcal{O} \mathcal{O} $\upsilon \upsilon \upsilon \upsilon$ UU \mathcal{O} \mathcal{O} \mathcal{O} \mathcal{O} \mathcal{O} UU \mathcal{O} \mathcal{O} \mathcal{O} $\overline{\mathcal{O}}$ \mathbb{U} $\overline{\mathcal{O}}$ \mathcal{O} 0 12 0 0 \mathcal{O} 12 12 0 U $\mathcal{O}:\mathcal{O}$ U U 0000 U \mathcal{O} U \mathcal{O} 12 12 U $\mathcal{O}:\mathcal{O}$ 12 0 0 0 0000000000000 \mathcal{O} \mathcal{U} $\upsilon \upsilon \upsilon \upsilon$ \mathcal{O} $\mathcal{O}\mathcal{O}$ 00 \mathcal{O} \mathcal{O} \mathcal{O} \mathcal{O} 00 00 UU UU \mathcal{U} U $\upsilon \upsilon \upsilon \upsilon$ \mathcal{O} $\upsilon \upsilon \upsilon \upsilon$ v v \mathcal{O} \mathcal{O} \mathcal{O} . UU \mathcal{U} U \mathcal{U} \mathcal{O} U $\upsilon \upsilon \upsilon$ U U \mathcal{O} \mathcal{O} \mathbb{T} $\mathbb{C}[\overline{\mathcal{O}}]$ $\cdot \upsilon$ U 11 19 U U U U 11 \mathcal{U} U U \mathcal{O} U U U 11 11 \mathcal{U} \mathcal{U} U U U U U U U U U U UU U U 11 U 00 00 0000 U U טטטט 00 0 U U 00 00 00 U U U





+ UNICATION



1 9.9.9.9 Я Я ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ S S S S S S S ্র ব্ - <u>5</u>-\$5:5 . ۲ হ 5: