The Traveling Observer Model

Multi-task Learning Through Spatial Variable Embeddings

Elliot Meyerson
Cognizant AI Labs

Risto Miikkulainen
UT Austin and Cognizant AI Labs

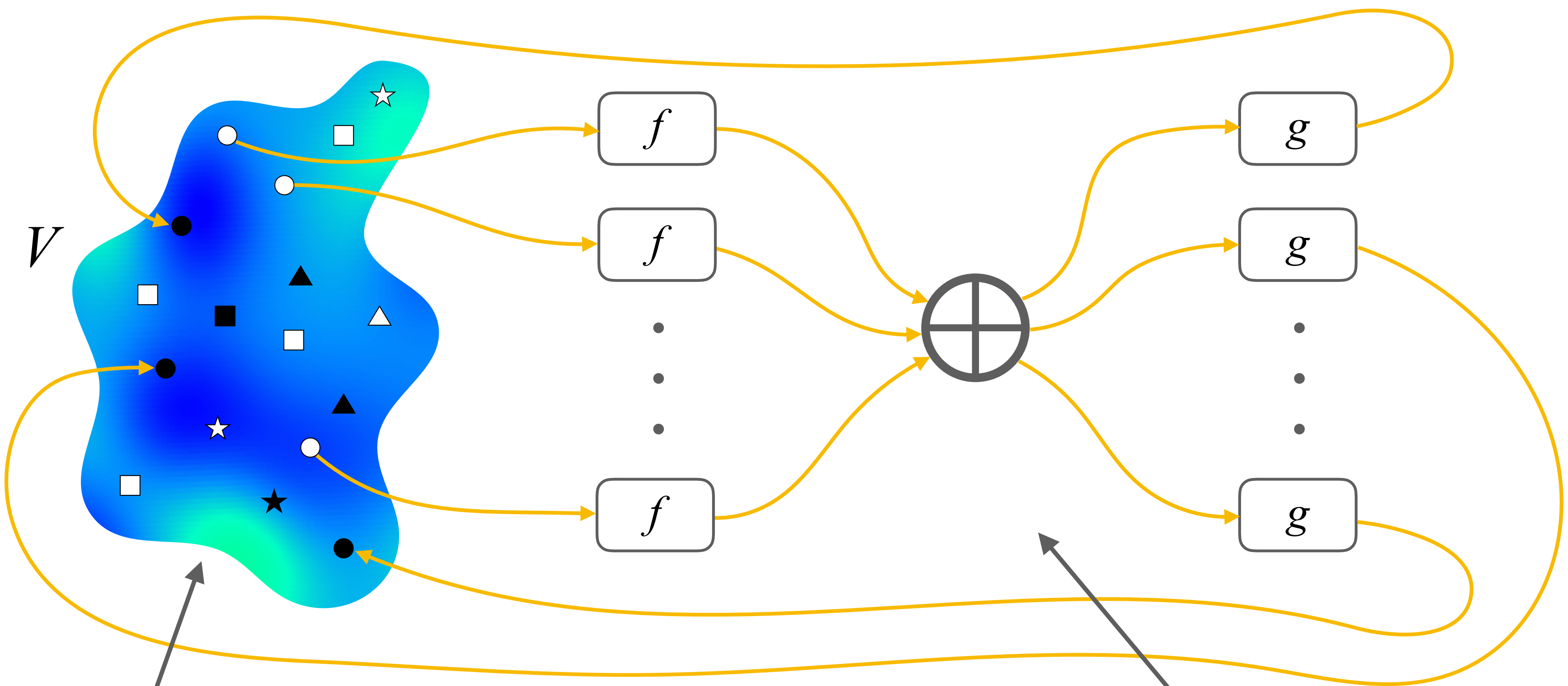
ICLR 2021

Measurements Organized in a Shared Space

(CBC, The Nature of Things)

- All human observations are embedded in the physical universe
- Multi-task learning community has avoided non-spatial disjoint tasks

The Traveling Observer Model (TOM)



Space where all variables are measured

Factor into encoder and decoder

Given observed locations and values, predict at unobserved location

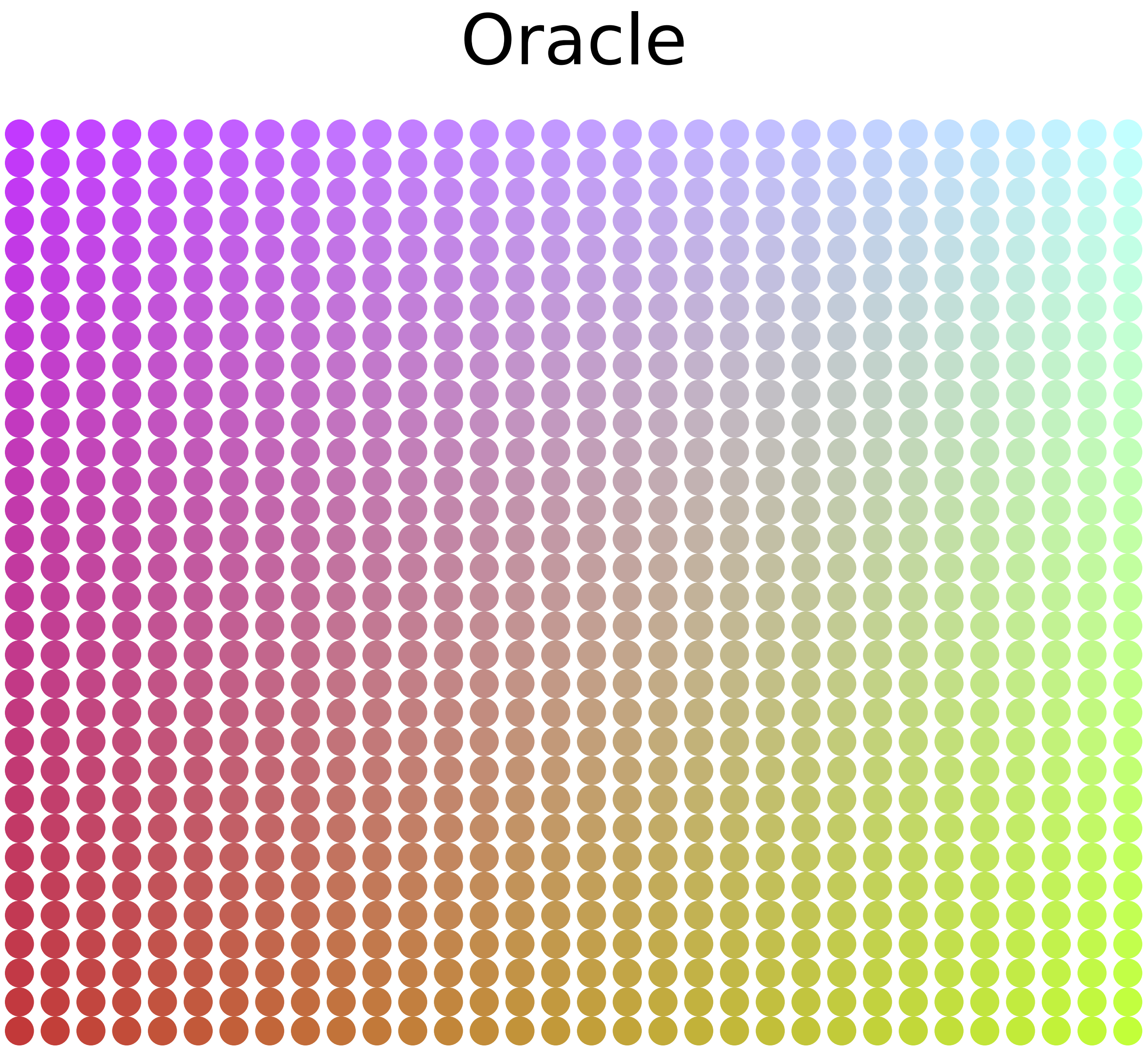
Implemented as ResNets with FiLM

$$\mathbb{E}[y_j \mid \mathbf{x}] = \Omega(\mathbf{x}, \{\mathbf{z}_i\}_{i=1}^n, \mathbf{z}_j) = g\Big(\sum_{i=1}^{n} f(x_i, \mathbf{z}_i),\ \mathbf{z}_j\Big)$$

Now Ω is a *universal predictor*

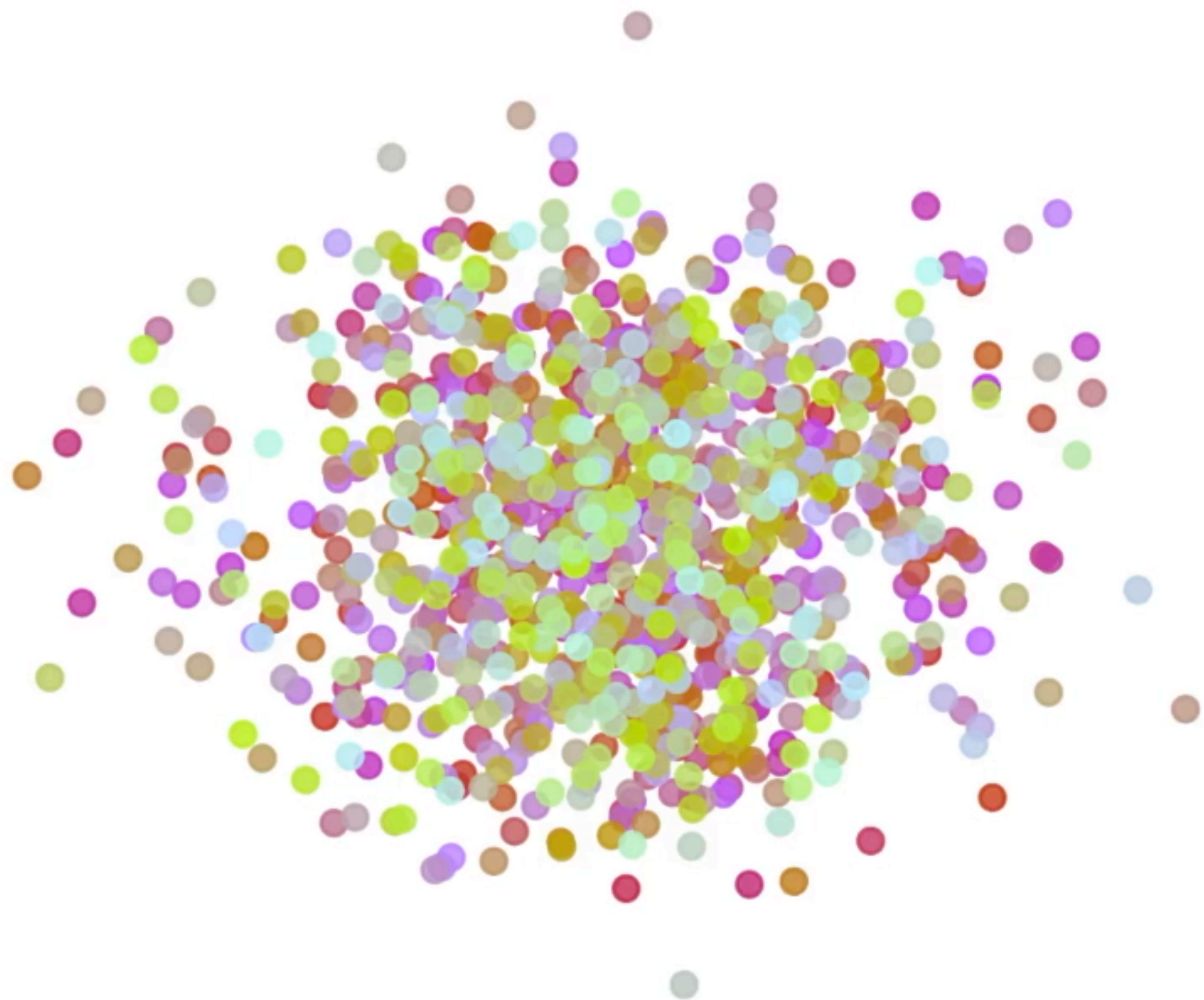
Variable embeddings $\mathbf{Z}$ can be learned by gradient descent

Discovering Space

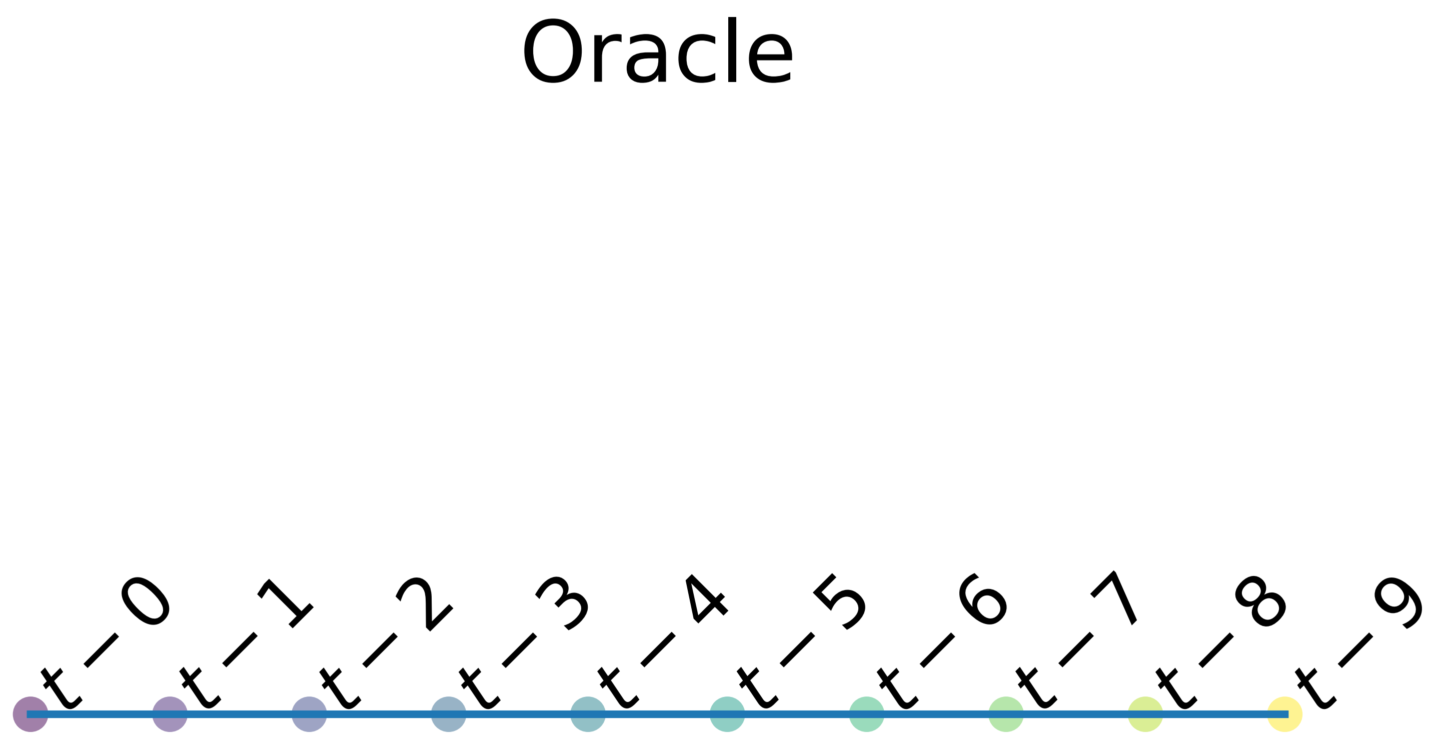


- CIFAR: Learning 2D variable embeddings of Pixel during image completion

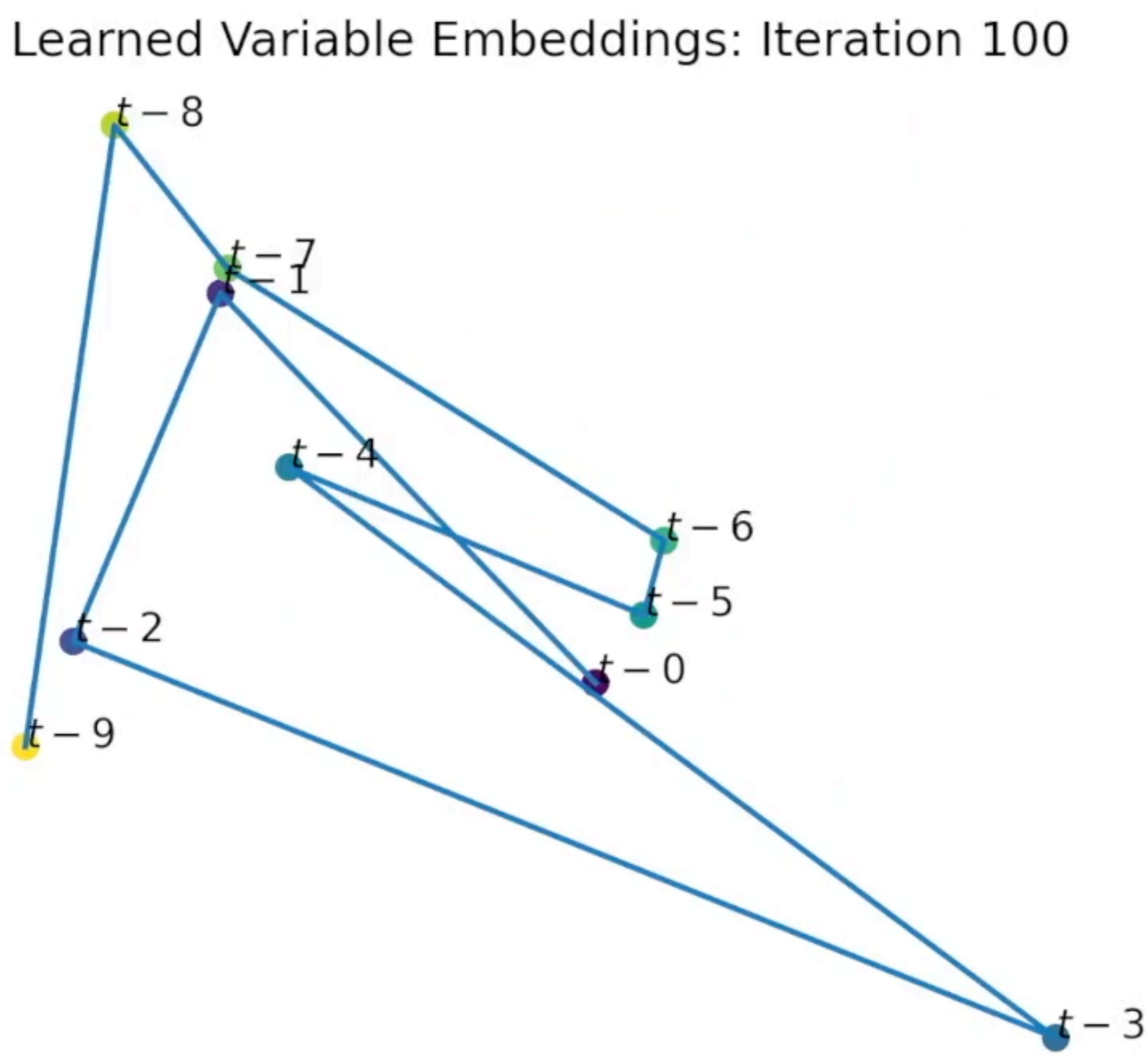
Learned Variable Embeddings: Iteration 0



Discovering Time



- Daily Temperature: Learning 2D variable embeddings of past days



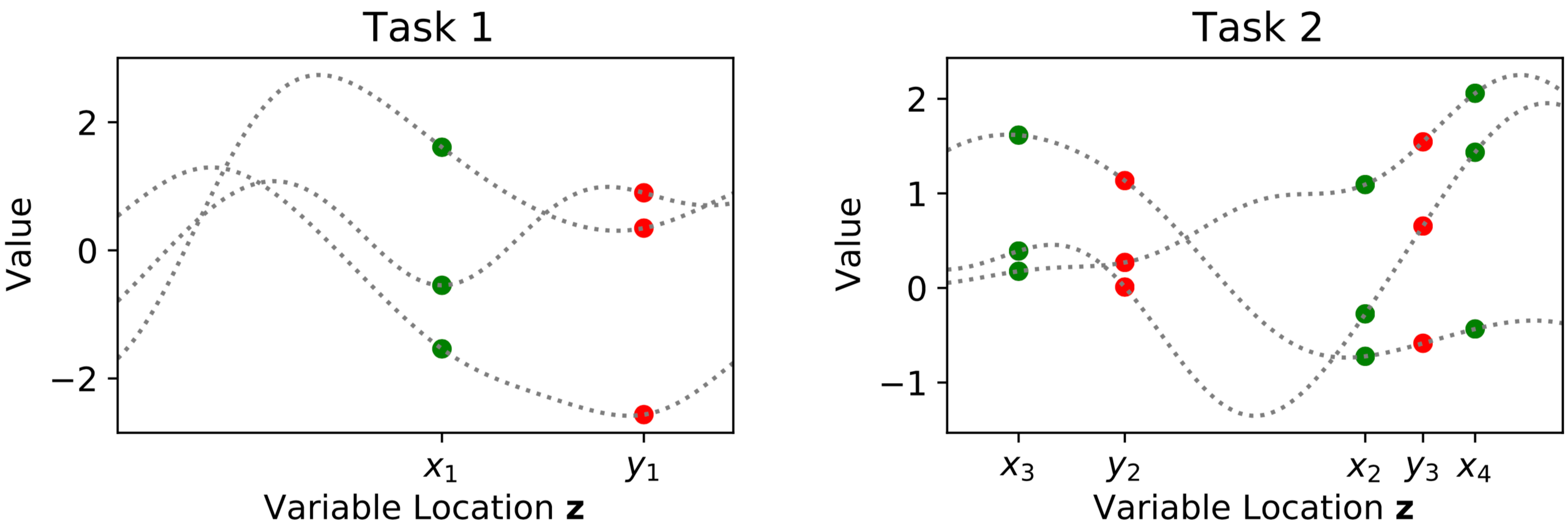
Learned Variable Embeddings: Iteration 100
$t-8$
$t-7$
$t-1$
$t-4$
$t-6$
$t-5$
$t-2$
$t-0$
$t-9$
$t-3$

Exploiting the Discovery of Space and Time

Variable Embeddings	Zero	Random	Learned	Oracle
CIFAR (Binary Cross-entropy)	0.662 ± 0.0000	0.660 ± 0.0007	**0.591** ± 0.0002	0.590 ± 0.0001
Daily Temperature (RMSE)	4.29 ± 0.002	4.27 ± 0.011	**3.32** ± 0.011	3.37 ± 0.005

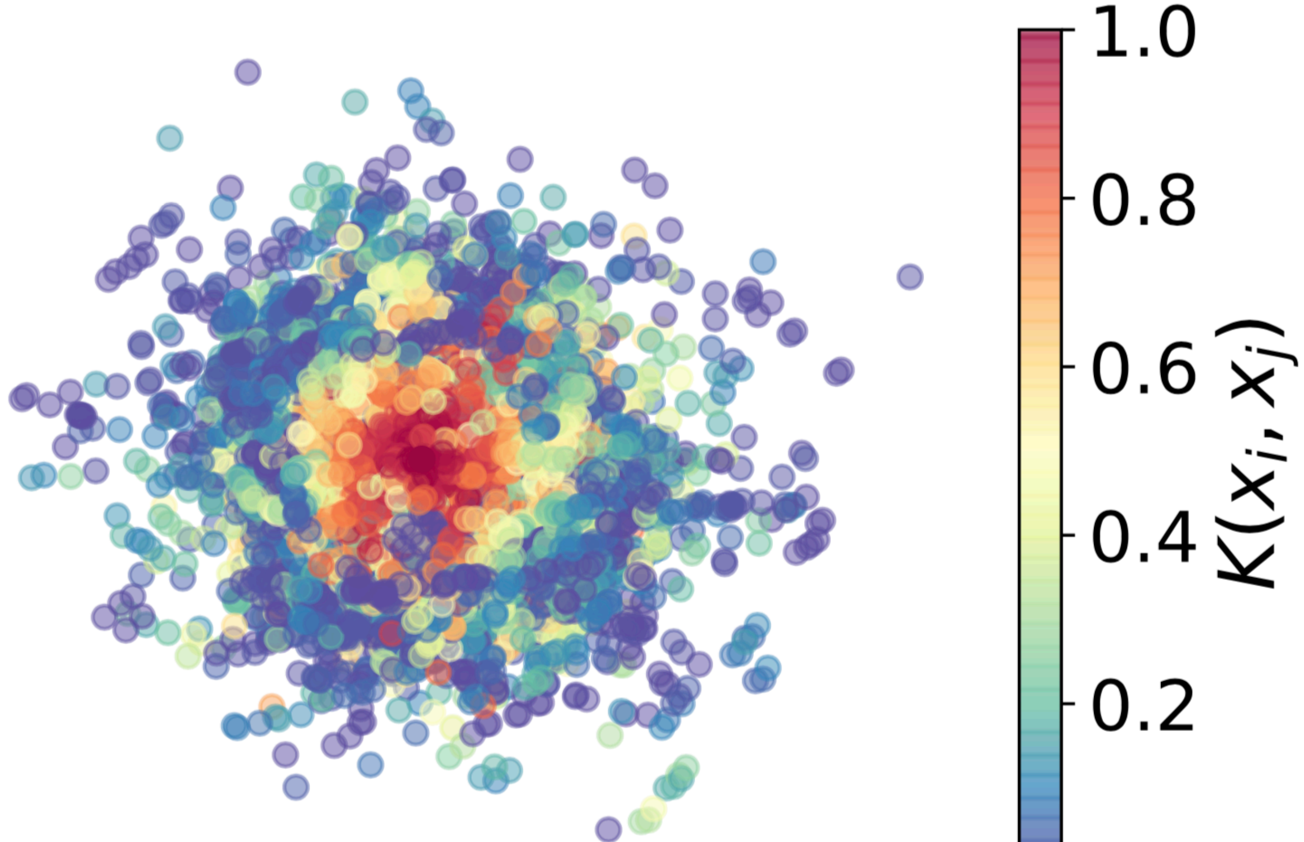
- Learned variable embeddings have performance on par with the Oracle.

Multi-task Regression: Transposed GP



- Tasks defined by variable locations in 1D Gaussian Process
- Nearby variables are more related; All variables affect all others
- 100 Tasks: 1 to 10 features; 1 to 10 targets

Multi-task Regression: Transposed GP

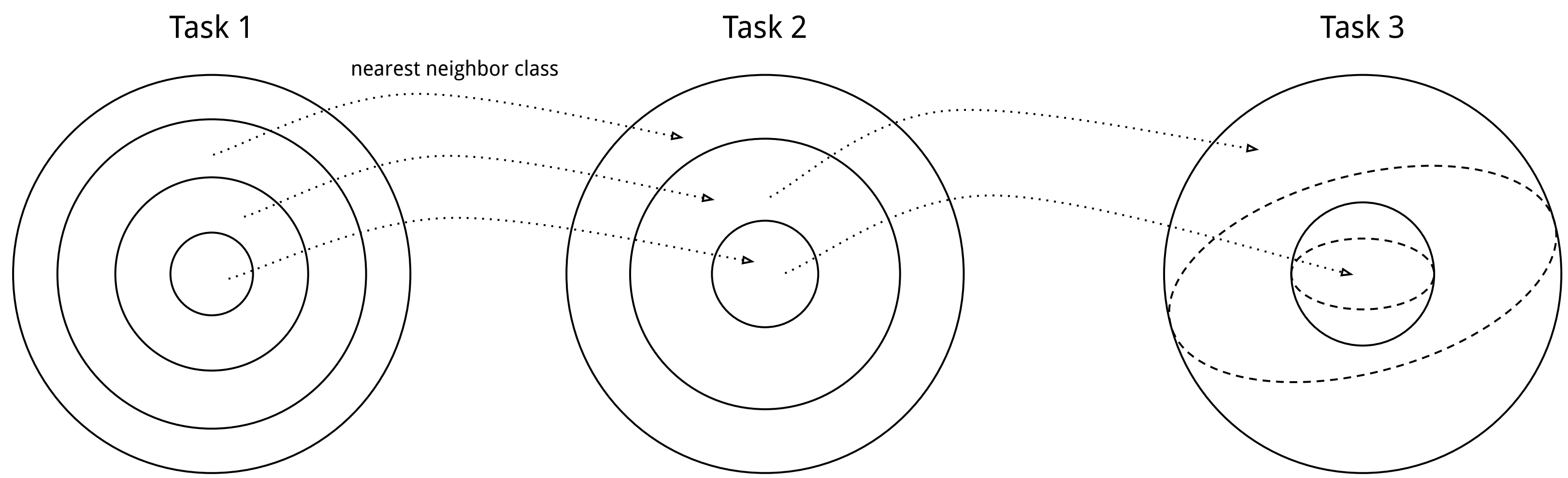


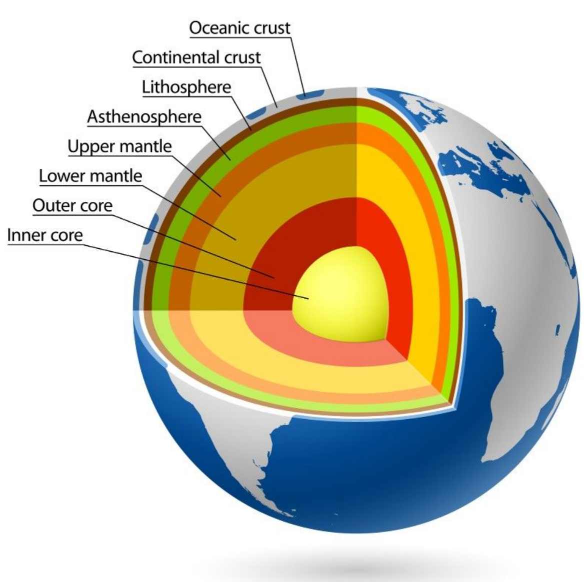
Kernel of learned variable embeddings

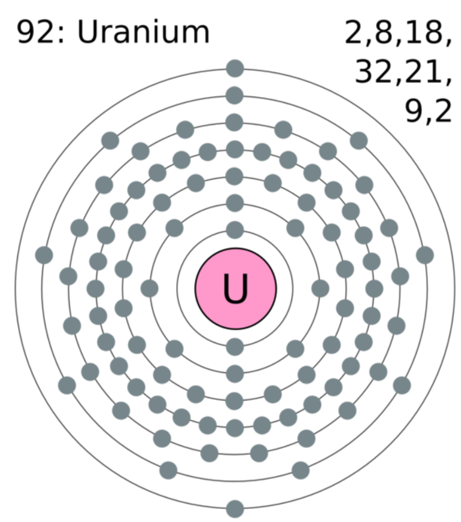
Method	Trans. GP (MSE)
DR-STL	0.373 ± 0.030
TOM-STL	0.552 ± 0.027
DR-MTL	0.397 ± 0.032
SLO	0.568 ± 0.028
TOM	**0.346** ± 0.031
Oracle	0.342 ± 0.026

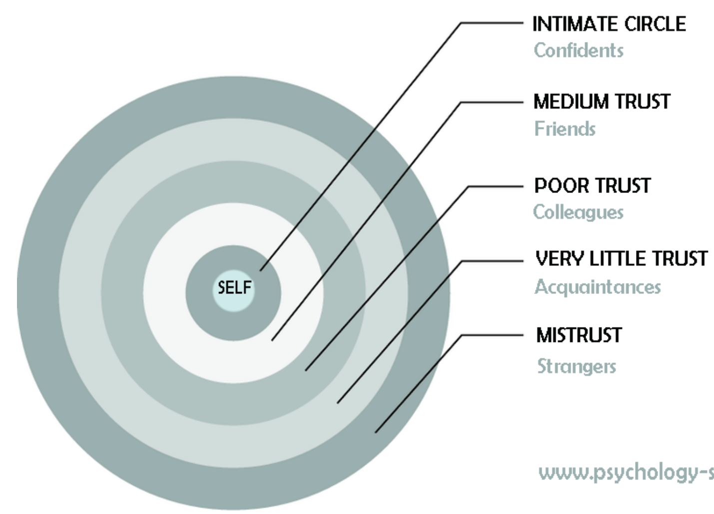
Test Performance

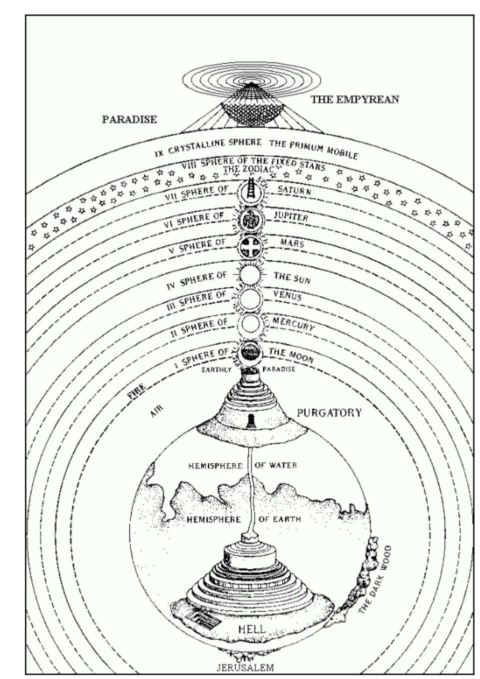
MTL Classification: Concentric Hyperspheres





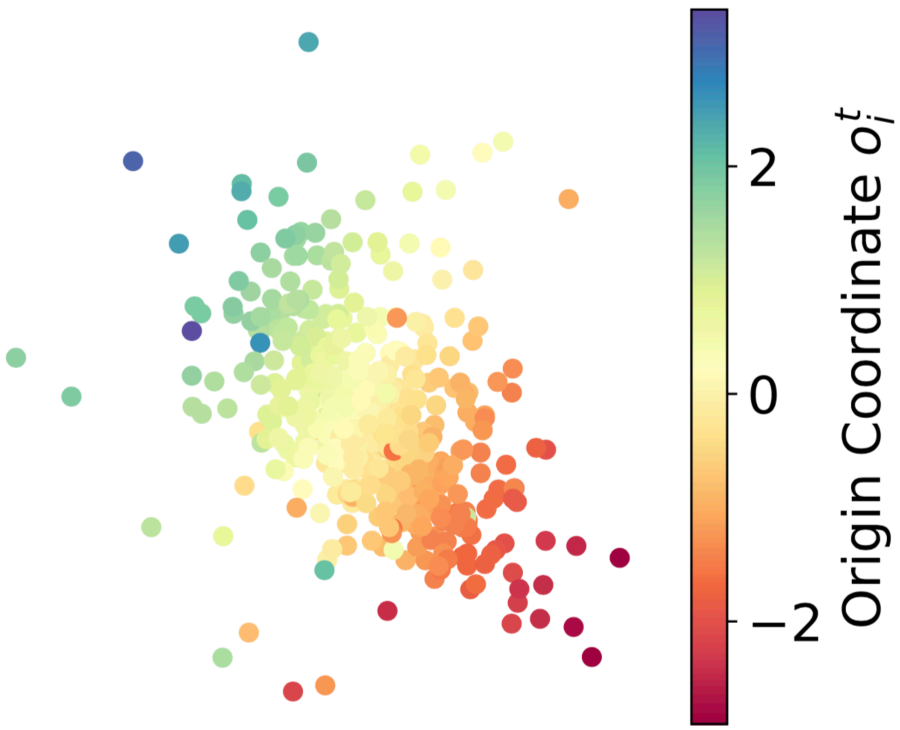


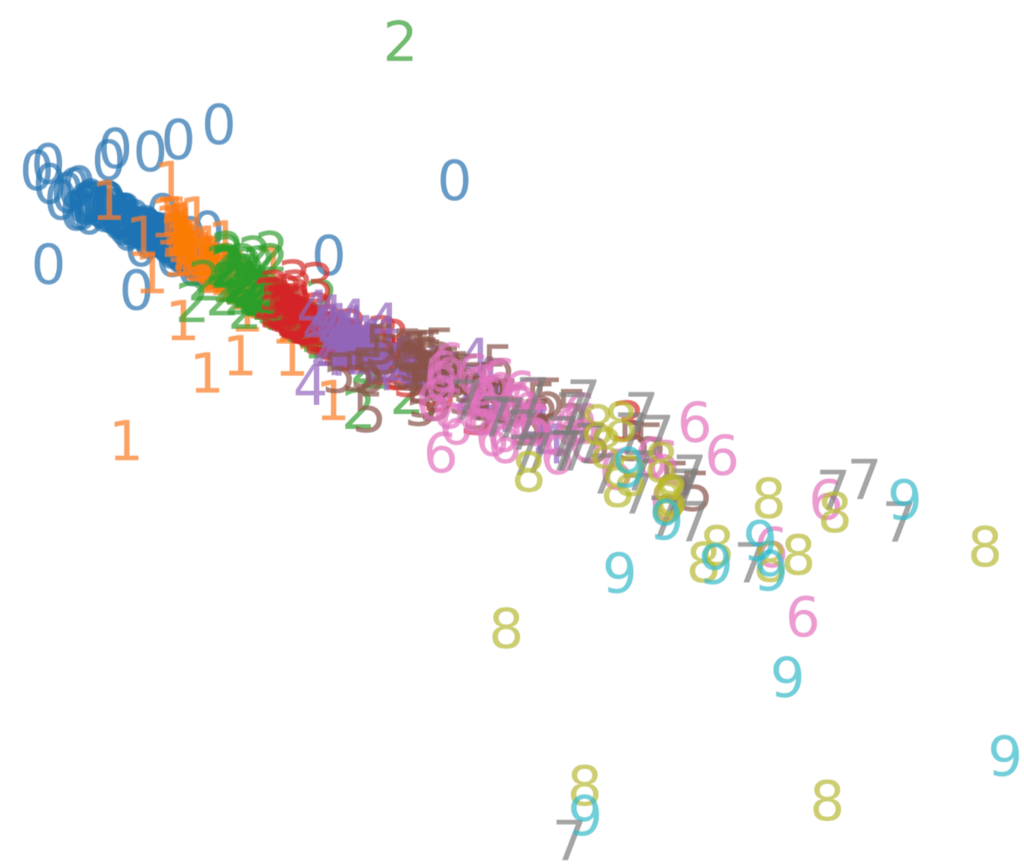




- Tasks defined by input dim, origin location and # annuli (classes)
- System must discover and exploit analogies to succeed.
- 90 Tasks: 1 to 10 features; 2 to 10 classes

MTL Classification: Concentric Hyperspheres





Learned variable embeddings of features and classes

Test Performance

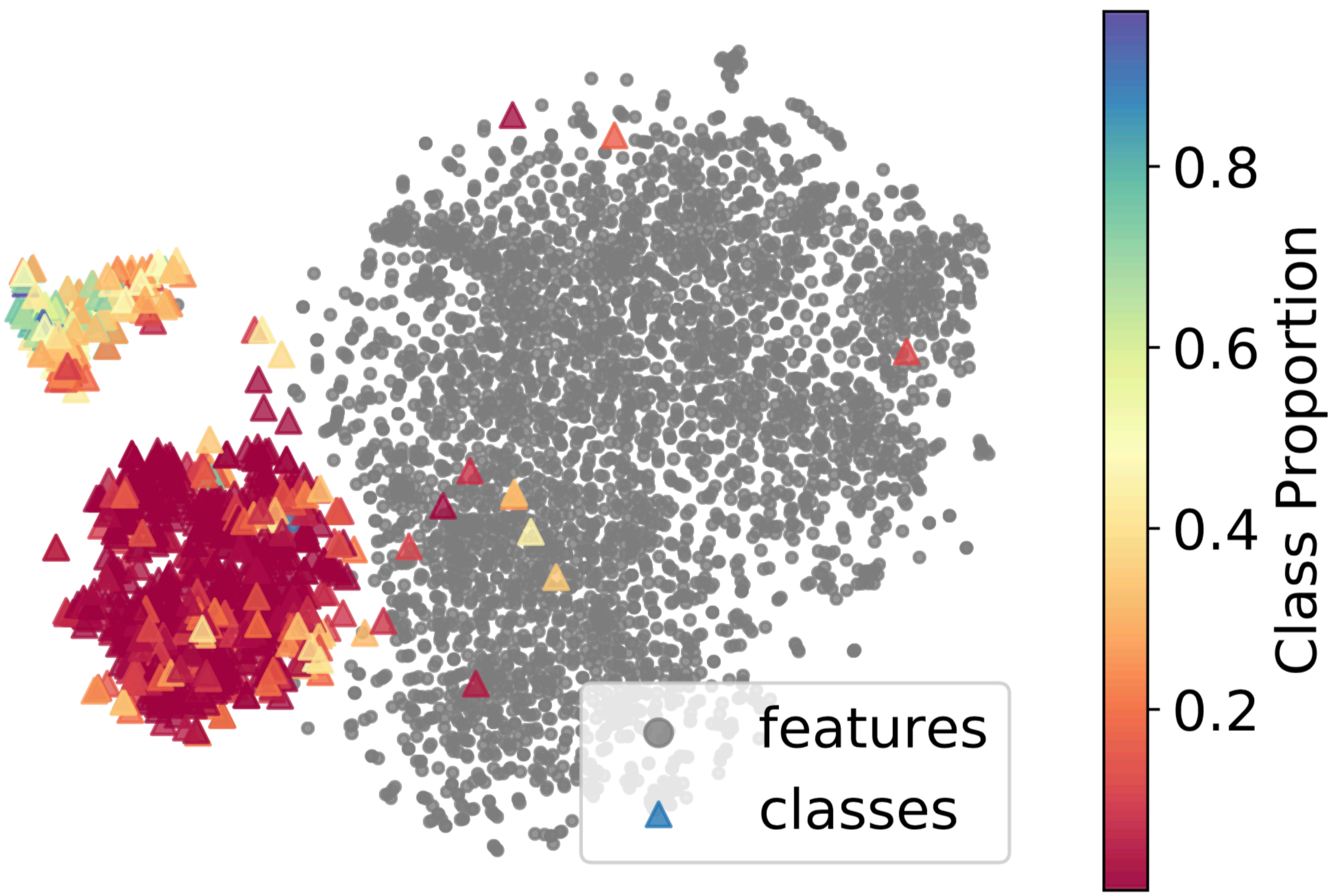
Method	C. Hypers. (Acc)
DR-STL	42.56 ± 1.69
TOM-STL	64.52 ± 1.83
DR-MTL	54.42 ± 1.92
SLO	53.26 ± 1.91
TOM	**92.90** ± 1.49
Oracle	99.96 ± 0.02

MTL over the distribution we care about: UCI-121

abalone	contrac	lenses	page-blocks	statlog-image
acute-inflammation	credit-approval	letter	parkinsons	statlog-landsat
acute-nephritis	cylinder-bands	libras	pendigits	statlog-shuttle
adult	dermatology	low-res-spect	pima	statlog-vehicle
annealing	echocardiogram	lung-cancer	pittsburg-bridges-MATERIAL	steel-plates
arrhythmia	ecoli	lymphography	pittsburg-bridges-REL-L	synthetic-control
audiology-std	energy-y1	magic	pittsburg-bridges-SPAN	teaching
balance-scale	energy-y2	mammographic	pittsburg-bridges-T-OR-D	thyroid
balloons	fertility	miniboone	pittsburg-bridges-TYPE	tic-tac-toe
bank	flags	molec-biol-promoter	planning	titanic
blood	glass	molec-biol-protein-second	plant-margin	trains
breast-cancer	haberman-survival	molec-biol-splice	plant-shape	twonorm
breast-cancer-wisc	hayes-roth	monks-1	plant-texture	vertebral-column-2clases
breast-cancer-wisc-diag	heart-cleveland	monks-2	post-operative	vertebral-column-3clases
breast-cancer-wisc-prog	heart-hungarian	monks-3	primary-tumor	wall-following
breast-tissue	heart-switzerland	mushroom	ringnorm	waveform
car	heart-va	musk-1	seeds	waveform-noise
cardiotocography-10clases	hepatitis	musk-2	semeion	wine
cardiotocography-3clases	hill-valley	nursery	soybean	wine-quality-red
chess-krvk	horse-colic	oocytes_merluccius_nucleus_4d	spambase	wine-quality-white
chess-krvkp	ilpd-indian-liver	oocytes_merluccius_states_2f	spect	yeast
congressional-voting	image-segmentation	oocytes_trisopterus_nucleus_2f	spectf	zoo
conn-bench-sonar-mines-rocks	ionosphere	oocytes_trisopterus_states_5b	statlog-australian-credit	
conn-bench-vowel-deterding	iris	optical	statlog-german-credit	
connect-4	led-display	ozone	statlog-heart	

- 121 Tasks: 3 to 262 features; 2 to 100 classes; 10 to 130,064 samples
- From diverse areas such as medicine, geology, engineering, botany, sociology, politics, game-playing, …

MTL over the distribution we care about: UCI-121



Learned variable embeddings

Method	Win %	Best %	Mean Rank	Norm. Acc.	Mean Acc.
ResNet	3.31 ±1.63	12.40 ±3.03	3.89 ±0.19	50.07 ±3.15	79.24 ±1.59
MS	4.96 ±1.98	14.88 ±3.28	3.35 ±0.19	60.11 ±3.00	80.11 ±1.48
BN	5.79 ±2.13	13.22 ±3.11	4.20 ±0.20	42.15 ±3.24	77.01 ±1.83
WN	7.44 ±2.40	10.74 ±2.84	4.05 ±0.20	45.87 ±3.11	77.43 ±1.74
HW	8.26 ±2.51	15.70 ±3.35	3.61 ±0.21	53.00 ±3.20	78.68 ±1.61
LN	9.92 ±2.73	16.53 ±3.40	3.45 ±0.20	56.73 ±3.03	79.85 ±1.53
SNN	13.22 ±3.09	21.49 ±3.78	2.78 ±0.19	65.29 ±2.84	81.39 ±1.35
TOM	**28.93** ±4.14	**34.71** ±4.36	**2.60** ±0.22	**70.72** ±3.02	**81.53** ±1.44
DR-STL	10.74 ±2.82	19.01 ±3.60	2.31 ±0.12	54.72 ±3.51	76.48 ±1.68
TOM-STL	7.44 ±2.40	16.53 ±3.40	2.72 ±0.13	35.21 ±3.72	68.18 ±2.26
DR-MTL	9.09 ±2.62	28.10 ±4.12	2.02 ±0.12	56.47 ±3.68	78.40 ±1.47
SLO	16.53 ±3.39	30.06 ±4.22	1.62 ±0.10	73.88 ±2.93	80.31 ±1.38
TOM	**32.23** ±4.27	**47.10** ±4.58	**1.34** ±0.13	**76.70** ±3.08	**81.53** ±1.44

Test Performance

- First time all tasks have been learned in a single joint model

Discussion

- ***Training regimes*: meta-learning, more sophisticated fine-tuning.**
- ***Theory*: desirable properties for distributions of variable embeddings?**
- ***Scale*: e.g., extensions to convolutions, attention, transformers.**
- ***Analogies*: TOM as an analogy engine, encoding task-agnostic concepts.**