Humans Learn Using Manifolds, Reluctantly

Bryan R. Gibson, Xiaojin Zhu, Timothy T. Rogers, Charles W. Kalish, Joseph Harrison

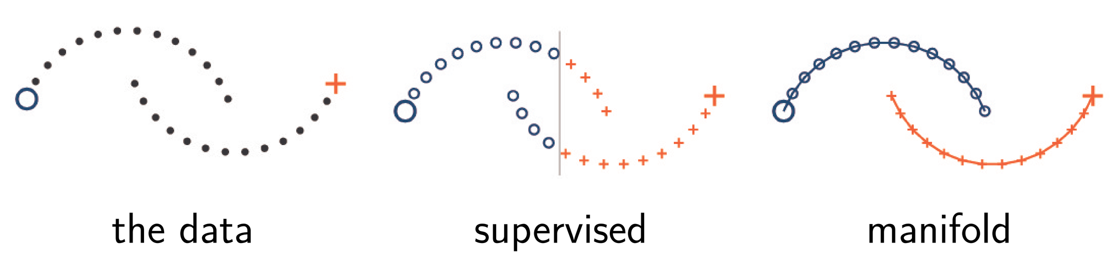
University of Wisconsin-Madison, USA

The Question

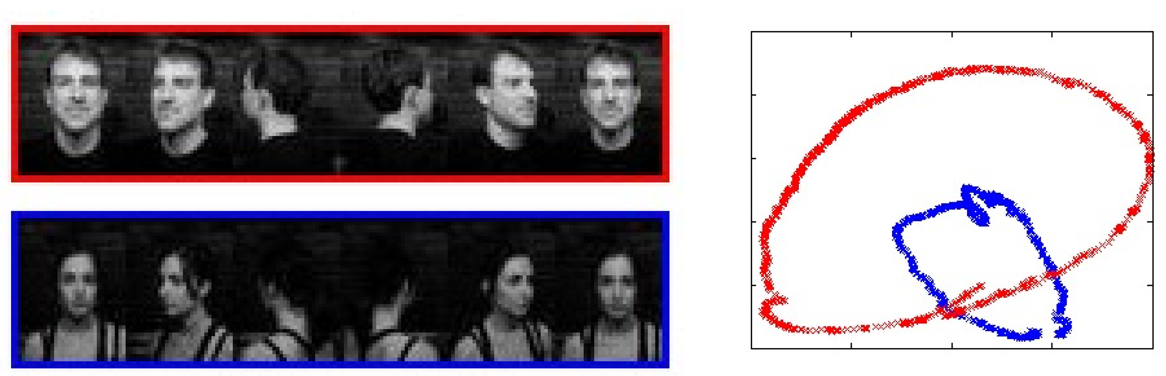
Under what conditions, if any, are people capable of manifold learning in a semi-supervised setting?

Manifold Learning

Consider a semi-supervised learning (SSL) task where the unlabeled examples are distributed along a lower dimensional plane, or manifold, in feature space. A manifold learner can take advantage of the underlying structure to propagate the label.



An Example: Face Recognition



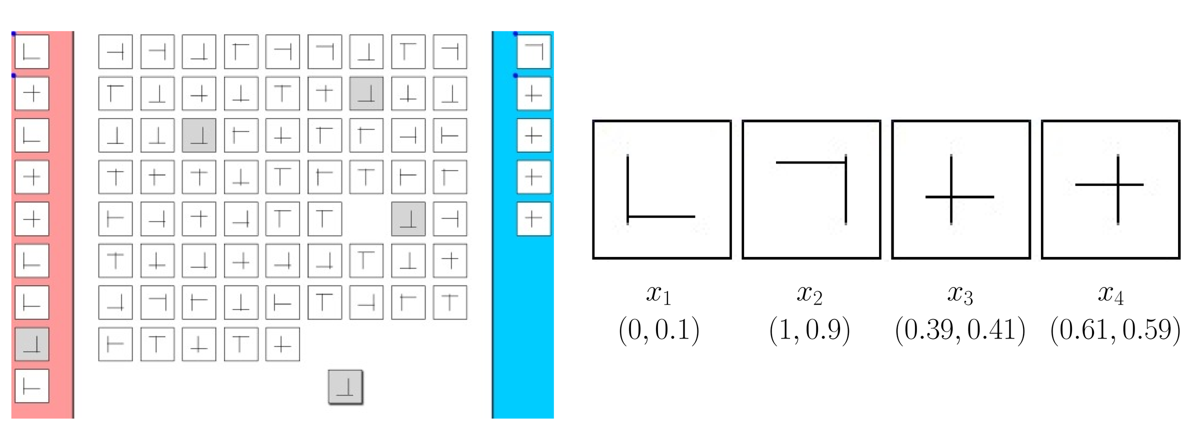
Projecting two series of 1000 images, each image 30x40 grayscale pixels or 1200 dimensions, down to 2 dimensions reveals the underlying manifolds.

The Experiment

139 participants in a SSL categorization task.

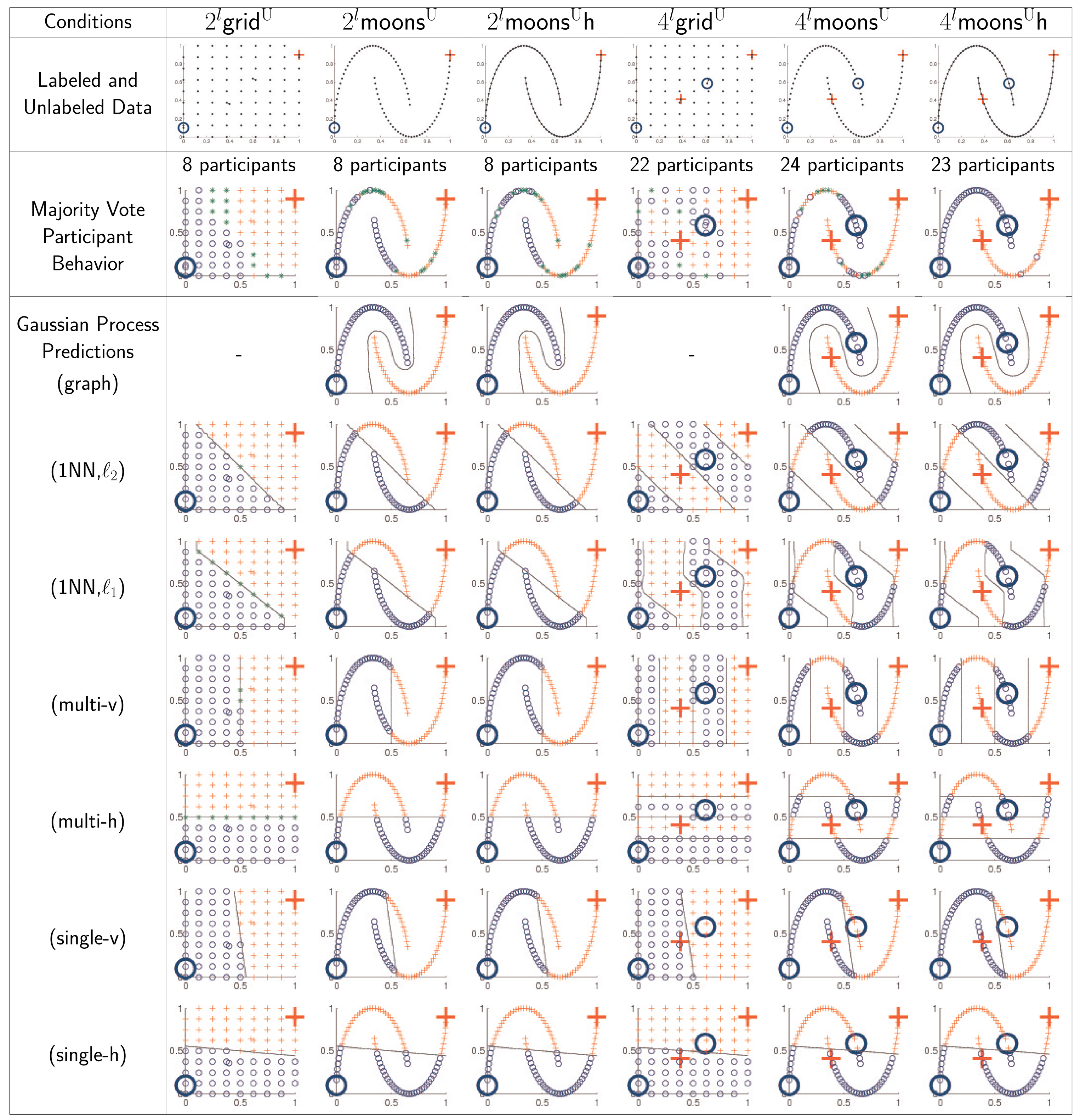
Manipulations:

- the number of labeled examples (2^l, 4^l),
- the underlying distribution used (grid^U, moons^U),
- whether graph neighbors were highlighted (h)



The task interface with highlighting and example stimuli.

Conditions, Behavior and Predictions



A set of simple classifiers simulated using a single Gaussian Process (GP) with different covariance functions, providing predictions for each condition.

Analysis

GP model accuracy in predicting human majority vote per condition.

	(graph)	(1NN,ℓ_2)	(1NN,ℓ_1)	(multi-v)	(multi-h)	(single-v)	(single-h)
2^lgrid^U	0.81	**0.94**	0.84	0.86	0.58	0.85	0.61
2^lmoons^U	0.47	**0.84**	0.62	0.74	0.42	0.79	0.45
2^lmoons^{U}h	0.50	**0.78**	0.56	0.76	0.36	0.76	0.39
4^lgrid^U	0.54	0.61	**0.64**	**0.64**	0.50	0.60	0.51
4^lmoons^U	0.64	0.62	0.60	**0.69**	0.47	0.38	0.45
4^lmoons^{U}h	**0.97**	0.76	0.54	0.64	0.31	0.65	0.26
4^lmoons^{U}h^R	**0.68**	0.63	0.44	0.56	0.40	0.59	0.42

Percentage of participants likely using each model per condition (argmax thresholded at 75%, 'other' indicates that no model was a good fit)

	(graph)	(1NN,ℓ_2)	(1NN,ℓ_1)	(multi-v)	(multi-h)	(single-v)	(single-h)	other
2^lgrid^U	0.12	0.00	0.12	**0.25**	**0.25**	0.12	0.00	0.12
2^lmoons^U	0.00	0.12	0.00	**0.25**	**0.25**	**0.25**	0.00	0.12
2^lmoons^{U}h	0.12	0.00	0.00	**0.38**	0.25	0.00	0.00	0.25
4^lgrid^U	0.00	0.05	0.09	0.00	0.00	0.18	0.09	**0.59**
4^lmoons^U	0.25	0.25	0.12	0.12	0.00	0.04	0.08	**0.38**
4^lmoons^{U}h	**0.39**	0.09	0.09	0.04	0.04	0.00	0.13	0.22
4^lmoons^{U}h^R	0.13	0.03	0.07	0	0	0.07	0.03	**0.67**

Observations

2^l → no manifold learning
2^l+ h → no manifold learning
4^l → no manifold learning
4^l+ h → manifold learning!

Learning or Following?

It is reasonable to ask, are participants learning the manifold or are they blindly following the highlighting?

1. 2^lmoons^{U}h→ no manifold learning, 4^lmoons^{U}h→ manifold learning, even though both had the same highlighting.
2. Participants perform 'leaps-of-faith'.

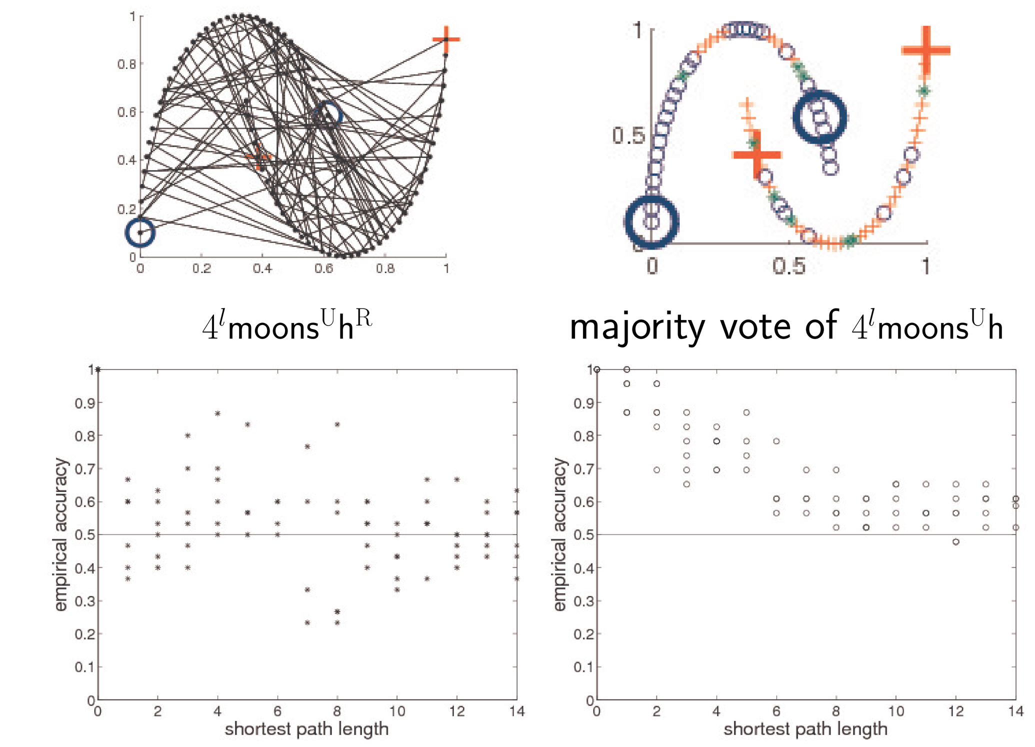
   A 'leap-of-faith' occurs when a participant classifies x as class y while all x's neighbors are either unlabeled or have labels other than y.

   4^lmoons^{U}h: average of 17 leaps-of-faith out of about 78 classifications.

   Blindly following the highlighting would yield zero leaps-of-faith.
3. Is the underlying manifold structure irrelevant?

   A new random manifold graph is created by taking a random permutation of the 4^lmoons^{U}h graph and mapping the i'th unlabeled point to the new permutation while keeping the adjacency matrix the same. 4^lmoons^{U}h^{R}and 4^lmoons^{U}hboth have two connected components with consistent labeled points.

   Blindly following means participants are more likely to classify unlabeled points the same as a labeled point the nearer the two are along the graph. This correlation should be the same for 4^lmoons^{U}h^R and 4^lmoons^{U}h, but it wasn't.



The underlying manifold is relevant. The correlation between label propagation and distance along the graph from a labeled point is not the same for a random graph (4^lmoons^{U}h^R) and manifold graph (4^lmoons^{U}h).

Evidence indicates participants are not blindly following the highlighting!

Model Selection

Using Bayesian model selection to explain the human behaviors, we can calculate the evidence $p(y_{1:l}|x_{1:l}, k)$ on labeled data for each kernel k used in the GP. The model selected is the one with the highest evidence.

The (graph) model is not the most likely model in 2^lmoons^{U}h while it is in 4^lmoons^{U}h, which matches participant behavior.

	2^lmoons^{U}h	4^lmoons^{U}h
(graph)	0.249	**0.0626**
(1NN,l2)	0.250	0.0591
(1NN,l2)	0.250	0.0625
(multi-v)	0.250	0.0625
(multi-h)	0.250	0.0625
(single-v)	0.249	0.0341
(single-h)	0.249	0.0342

The Punchline

People can learn the half-moons dataset, if we give them 4 (not just 2) labeled points and give them clues about the graph.

Research supported in part by NSF IIS-0916038, NSF IIS-0953219, NSF DRM/DLS-0745423, and AFOSR FA9550-09-1-0313.

NIPS 2010 Vancouver, BC {bgibson,jerryzhu}@cs.wisc.edu