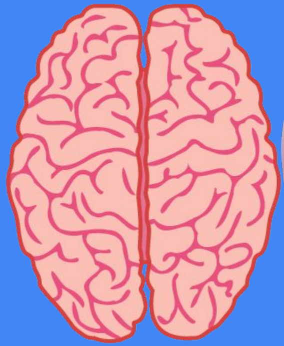
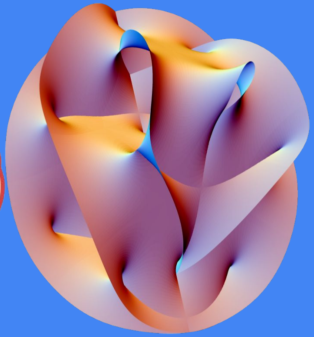
 

The Many Folds of Cognition

A Topological Perspective on Form Learning

Anshul Bawa
Sahil Loomba

10 Nov 2015

Questions around Cognition

- How do we learn inductively?
- Given time and resource constraints, which heuristics do we use for “efficient computation”?
- Are different percepts organised semantically in our cognitive system?

Questions around Cognition

- How do we learn inductively?
  - **FULL BAYESIAN LEARNING**
- Given time and resource constraints, which heuristics do we use for "efficient computation"?
  - **FEATURE COMPACTION**
- Are different percepts organised semantically in our cognitive system?
  - **FORM LEARNING**

Cognitive Heuristics

Deviations from perfect rationality

- Representativeness Heuristic
- Availability Heuristic

Compact Knowledge Representation

Memory - Percepts to concepts

Semantic gist-of-things

Information aggregation

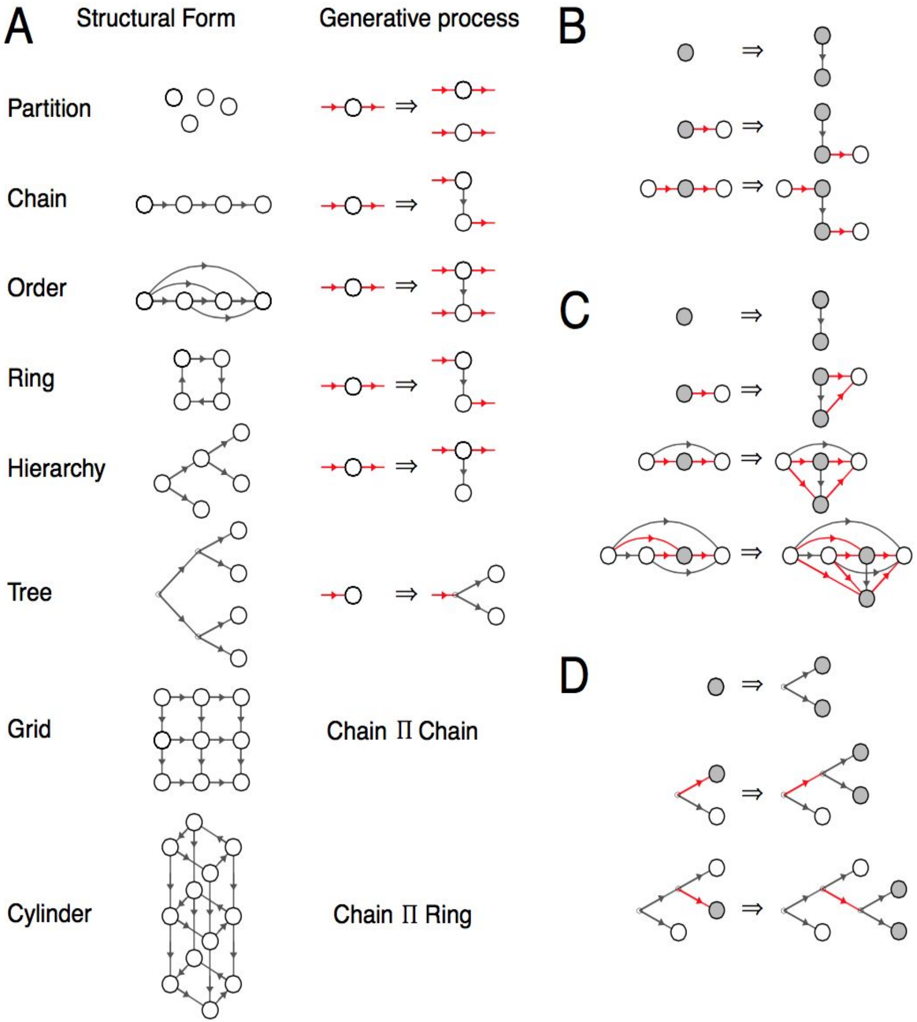
Form Learning
[Kemp Tenenbaum '08]

Unified framework - Graph grammar

Dataset - Animal features

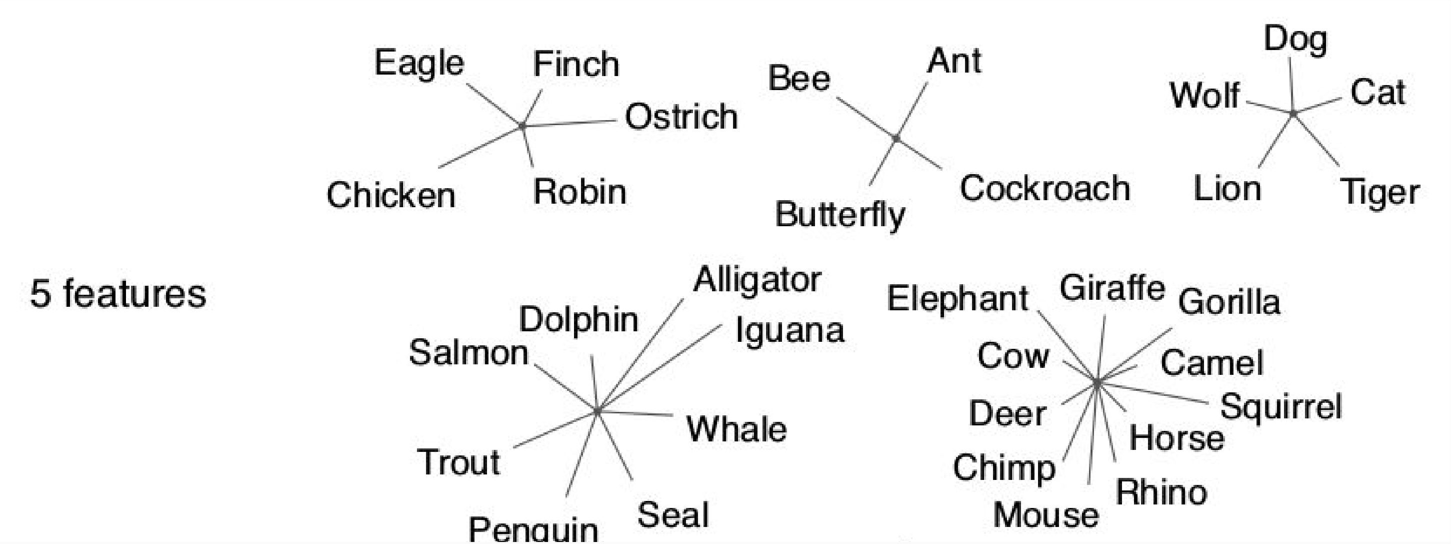
Learning on raw data

$$P(S, F|D) \propto P(D|S)P(S|F)P(F).$$



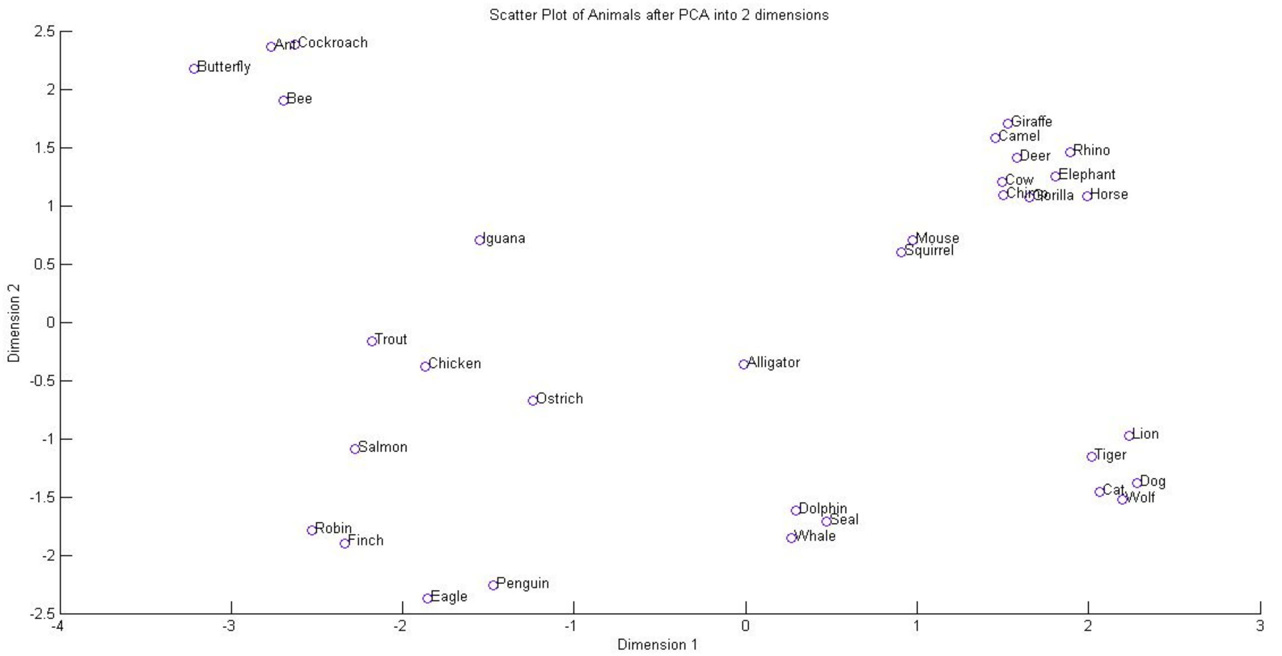
Dimensionality Reduction

Naive - Subset of Features



Dimensionality Reduction
Linear - Principal Component Analysis

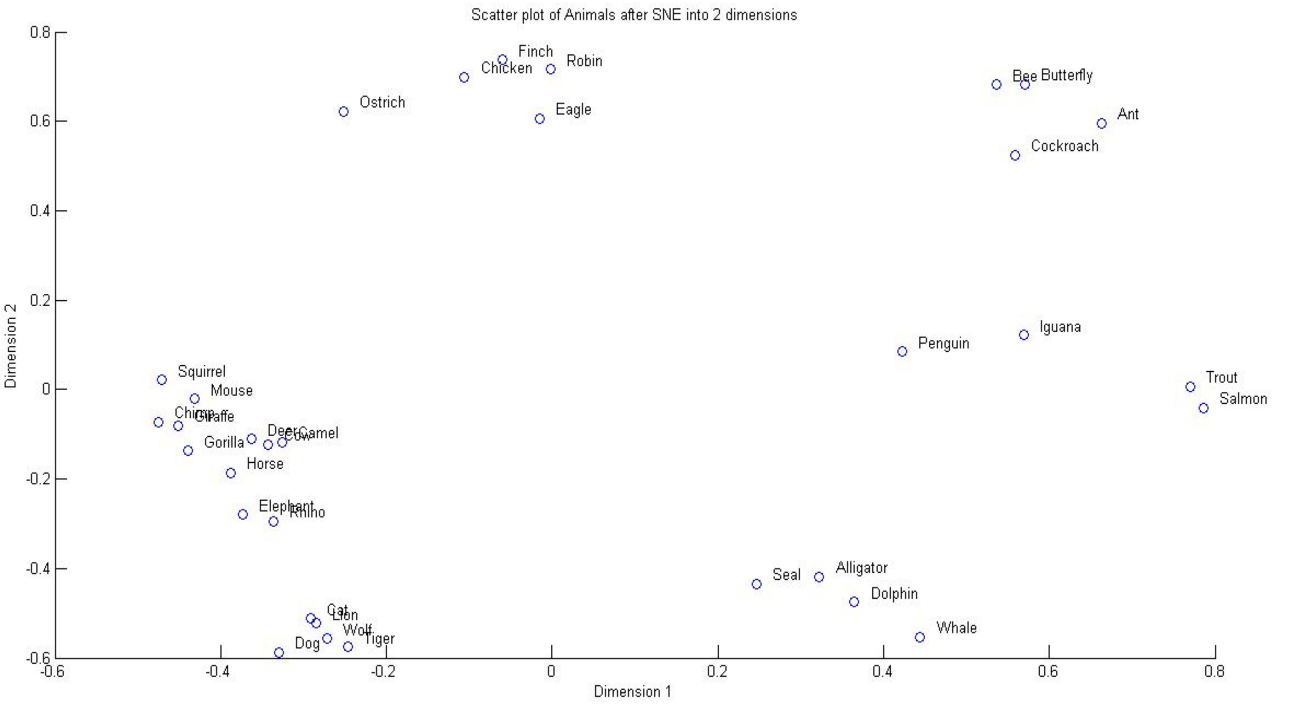
Resolves data into orthogonal dimensions using purely covariance



Dimensionality Reduction
Manifold - Stochastic Neighbour Embedding

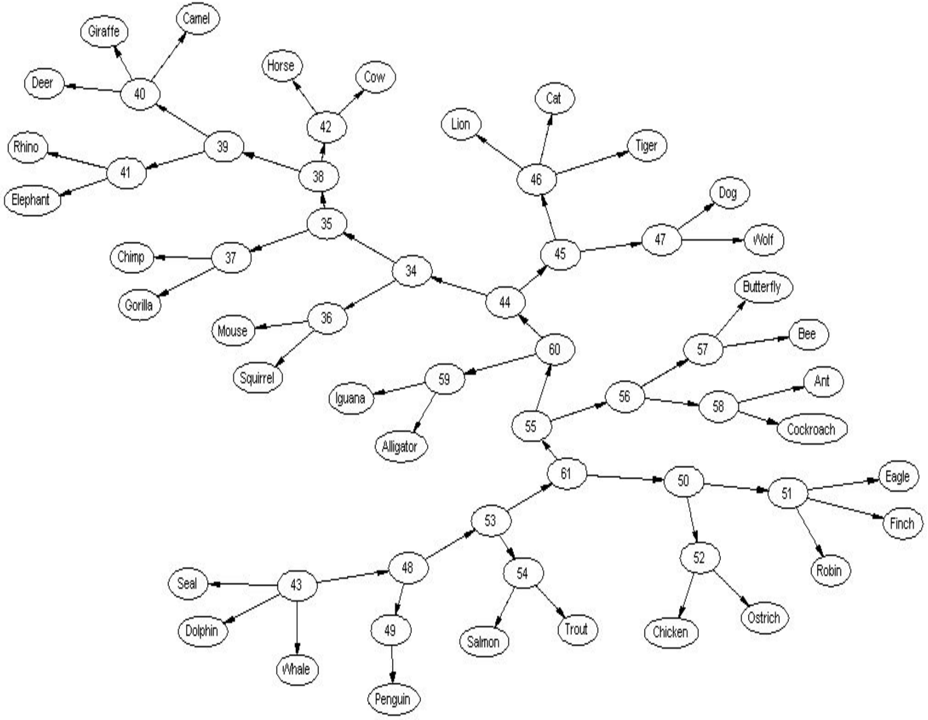
Centers a Gaussian at every point in high-dimensional space

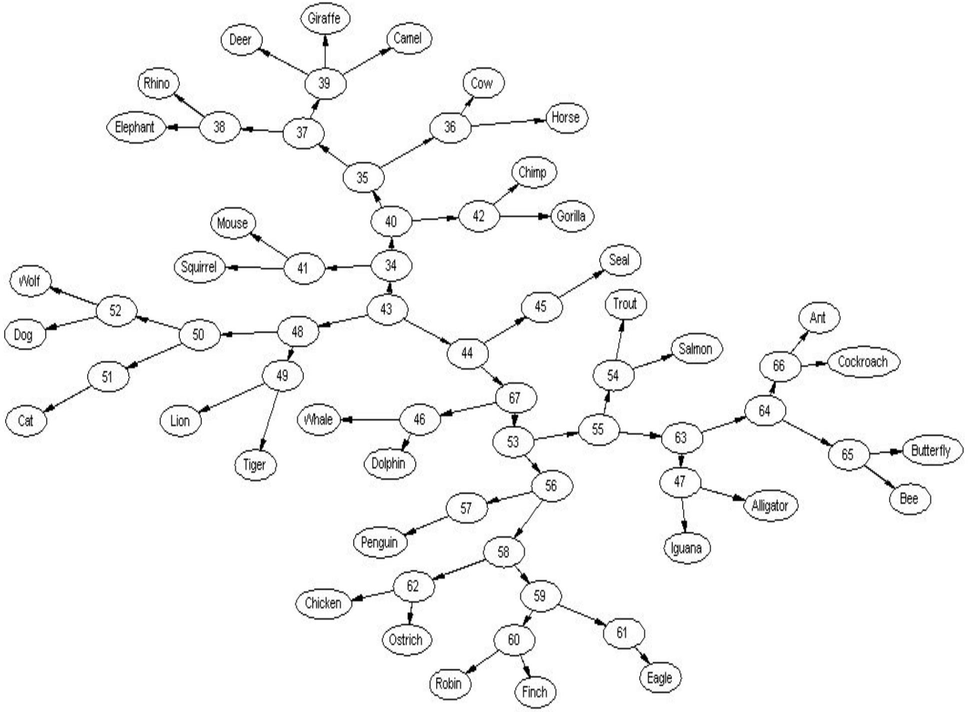
Preserves density map in lower-dimensional space



Why Manifolds Work

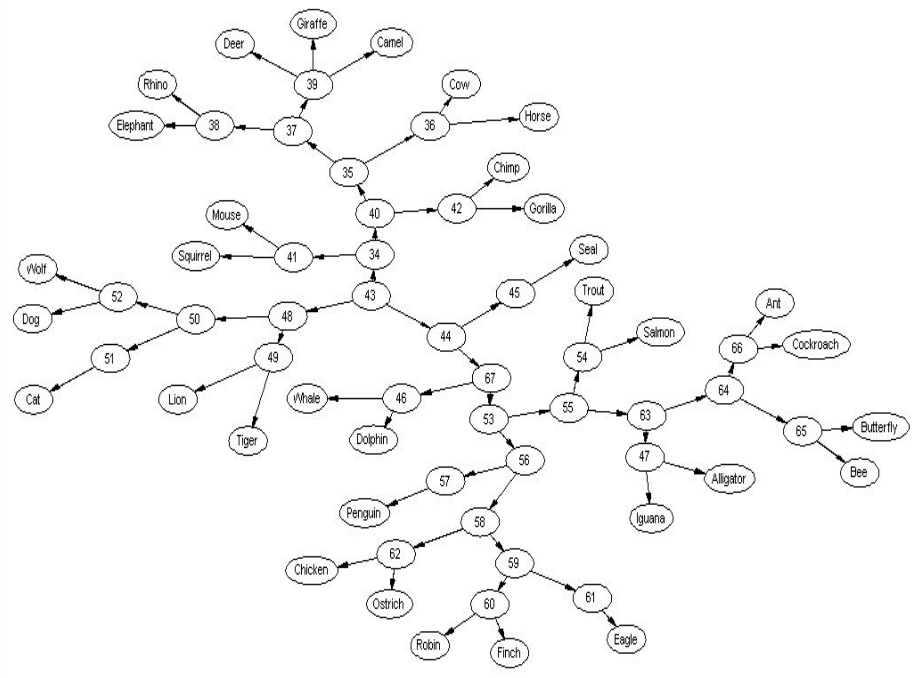
Tree Structure Likelihood: SNE over Ground





Why Manifolds Work

Tree Similarities: SNE over PCA

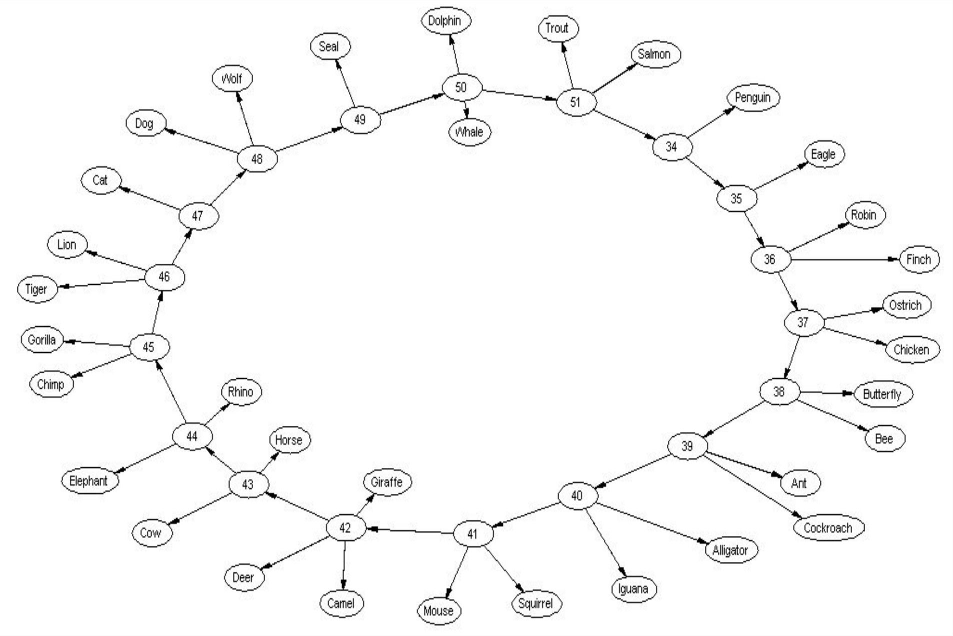


Dimensionality Reduction Method	Similarity of tree to true tree (lower is better)
PCA-2	0.9303
PCA-4	0.9327
PCA-8	0.9396
SNE-2	0.9280
SNE-4	0.9249
SNE-8	0.9063

Why Manifolds Work

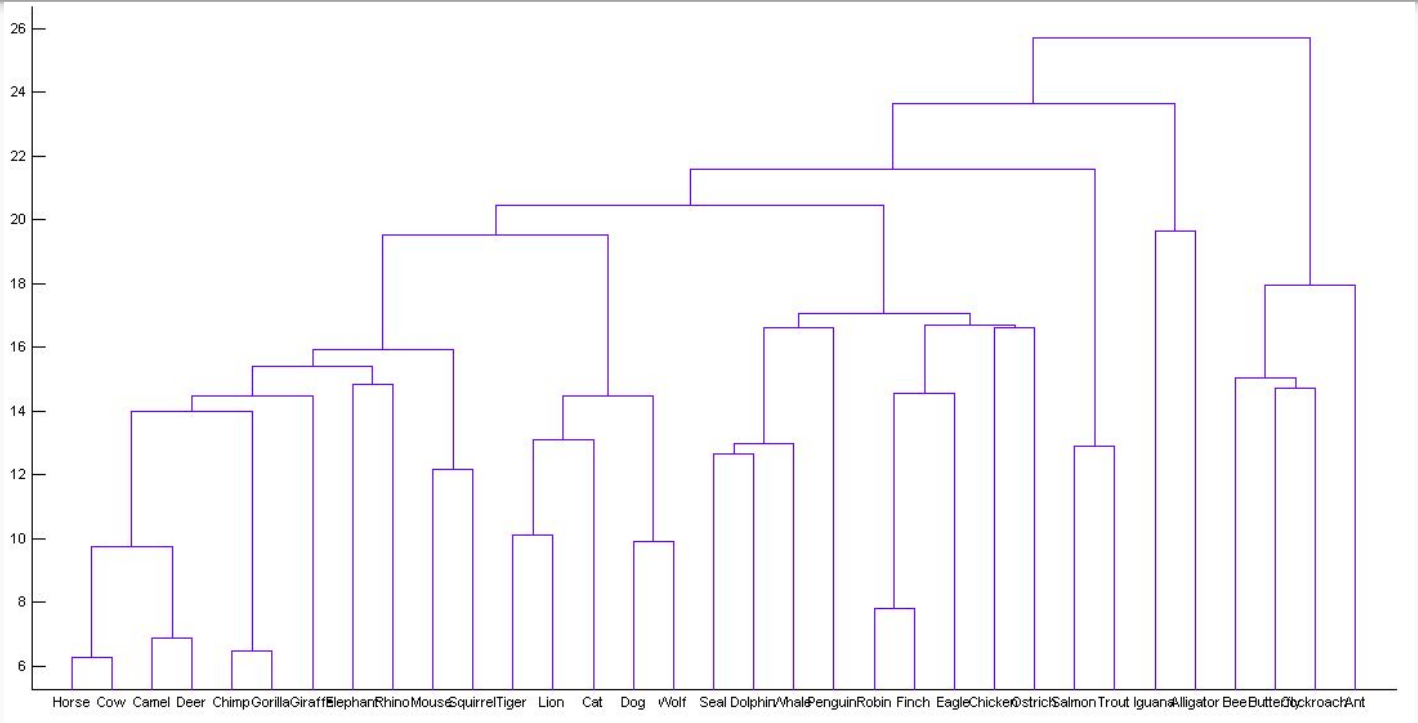
Ring (False) Learning: SNE over Ground

Dataset	Log of relative likelihood of tree w.r.t ring (higher is better)
Ground	-2.7
SNE-4	1.7



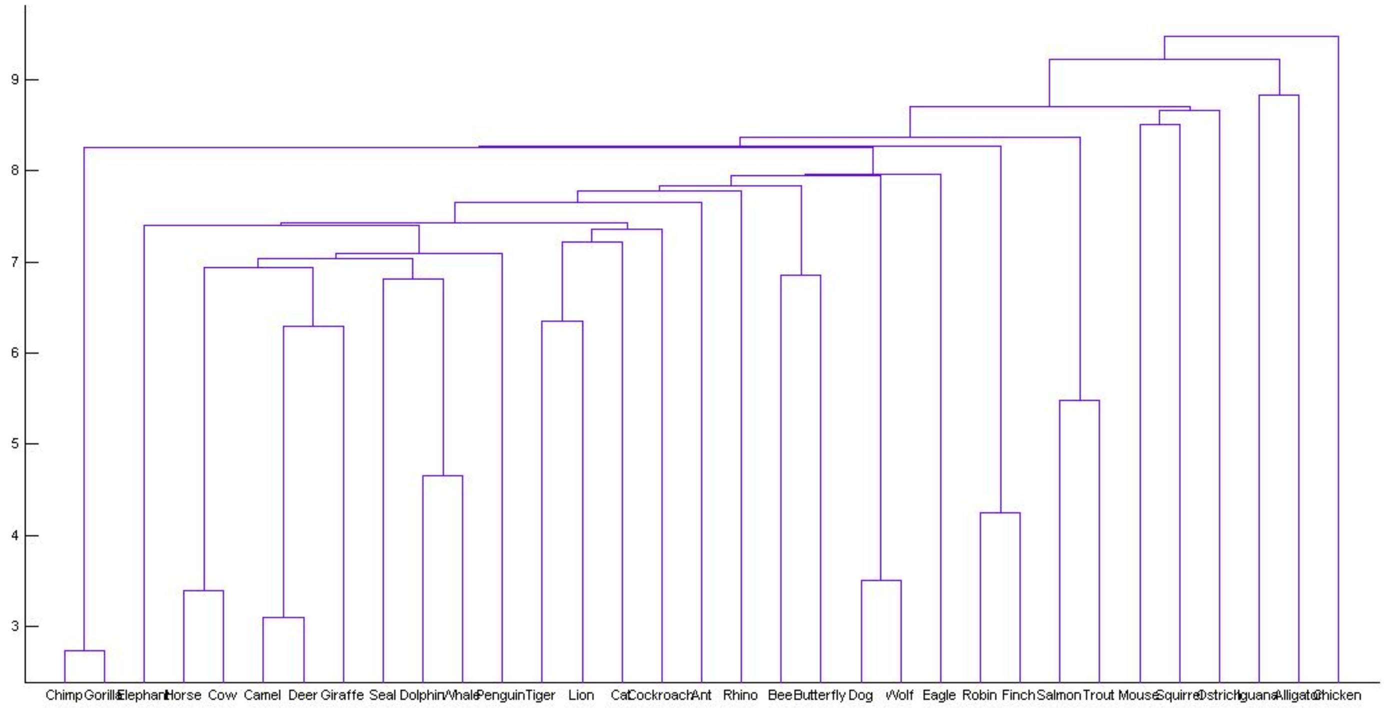
Why Manifolds Work

Hierarchical Agglomerative Clustering - Ground



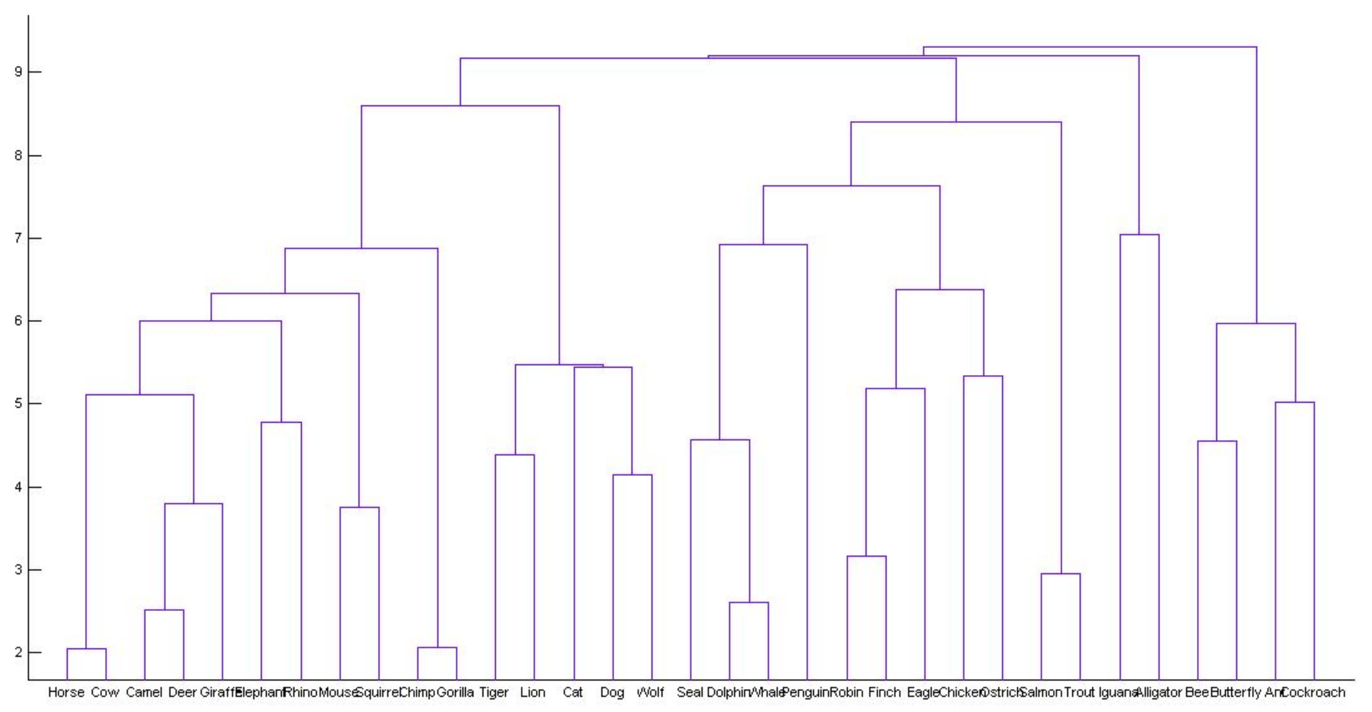
Why Manifolds Work

Hierarchical Agglomerative Clustering - PCA



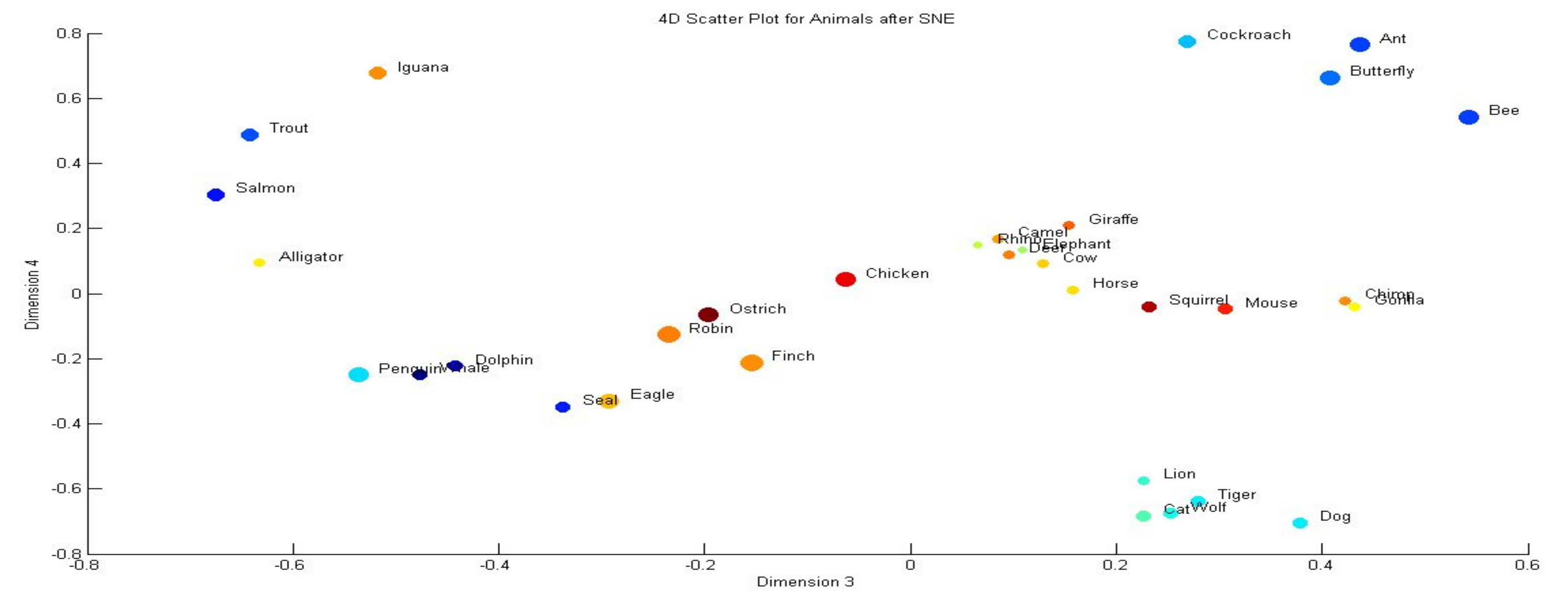
Why Manifolds Work

Hierarchical Agglomerative Clustering - SNE



Why Manifolds Work

Representativeness Heuristic



Why Manifolds Work

Discriminating Features and the Availability Heuristic

S.No	List of discriminating features	Type of feature
1	has a large brain	Anatomical (visible)
2	has 6 legs	Anatomical (visible)
3	has a nose	Anatomical (visible)
4	has paws	Anatomical (visible)
5	has antennae	Anatomical (visible)
6	is long	Anatomical (visible)
7	is large	Anatomical (visible)
8	has tusks	Anatomical (visible)
9	is slender	Anatomical (visible)
10	has horns	Anatomical (visible)
11	has hooves	Anatomical (visible)
12	is poisonous	Anatomical
13	is soft	Anatomical (visible)
14	is black	Anatomical (visible)
15	is a rodent	Anatomical (visible)

18	is an insect	Anatomical (visible)
19	is scaly	Anatomical (visible)
20	is furry	Anatomical (visible)
21	has flippers	Anatomical (visible)
22	is colorful	Anatomical (visible)
23	is a canine	Anatomical (visible)
23	is strong	Behavioural
25	howls	Behavioural
26	travels in groups	Behavioural
27	is dangerous	Behavioural
28	digs holes	Behavioural
29	eats grass	Eating habits
30	eats leaves	Eating habits
31	eats bugs	Eating habits
32	eats fish	Eating habits
33	lives in lakes	Habitat

Why Manifolds Work

Discriminating Features and the Availability Heuristic

Fraction of total	All features	Discriminating features
Anatomical	0.57	0.64
Visible anatomical (as fraction of anatomical)	0.83	0.96
Behavioural	0.24	0.13

Tying it all together
Manifolds, Typicality, Bayesian cognition, Form learning

Tradeoff : Generality of gist-extraction vs. Generality of graph grammars.

Model <- Parameters <- Hyperparameters <- **Gist** <- Features <- Data

Typicality not only in Learning, but in Inference

Modelling not only what we do right, but also what we do wrong

On Inductive Learning

From ML, to MAP, to Full Bayesian Learning

Maximum Likelihood:

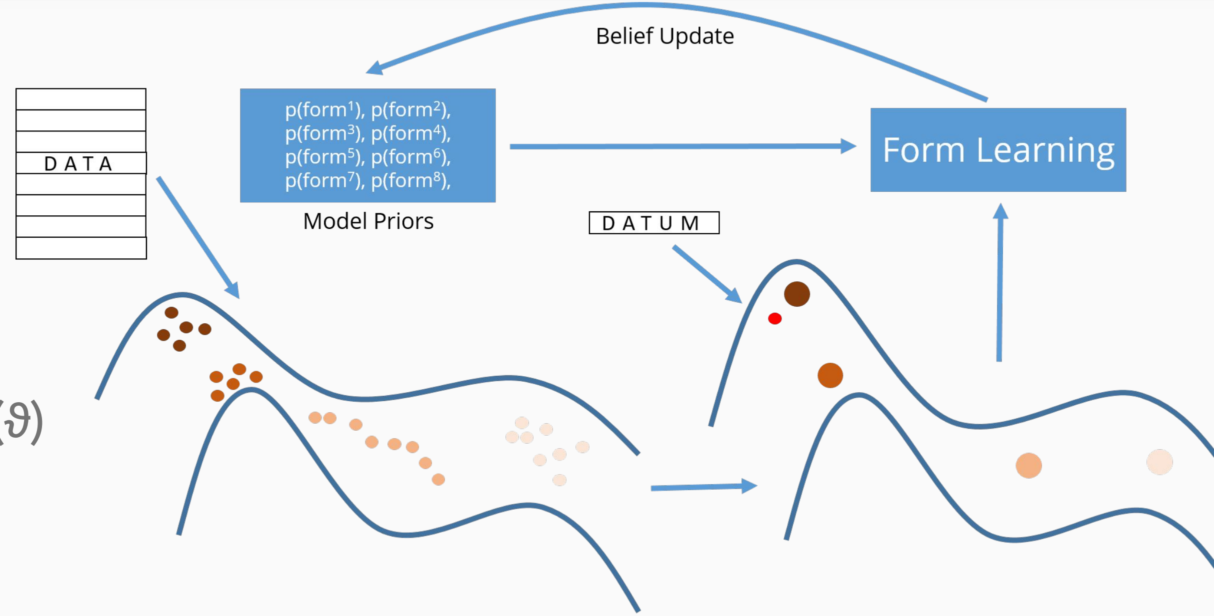
max P(X/ϑ)

Maximum a posteriori:

max P(ϑ/X) = max P(X/ϑ).P(ϑ)

Full **Bayesian** Learning:

mean P(X/ϑ).P(ϑ)



In Conclusion

“Indeed, the human mind appeared to suffer from a crippling need to fabricate in the absence of concrete proof.”

- J. R. Ward

Thank You