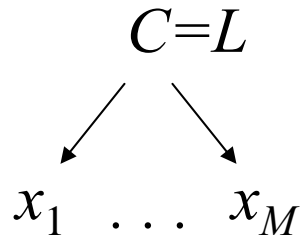


Anderson (1991): Rational Analysis

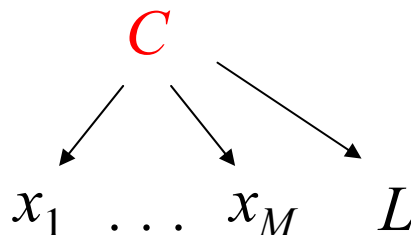
- Goal of categorization: feature prediction for biological species (or similar natural kinds).
- Structure of biological species: feature clusters
 - Some animals fly, have wings, and live in trees.
 - Some animals swim, have fins, and live in water.
 - Some animals walk, have legs, and live on earth.
 - No animals fly, have fins, and live on earth
- Observing a few features often allows us to infer many others:
 - e.g., observe an animal that has wings

Model architecture

- Prototype model:
 - Lexical labels identified with concepts



- Anderson model:
 - Lexical labels just another feature to predict.



Concept membership C
unobserved.

Image removed due to
copyright considerations.

Two interesting issues

- Applying an infinite mixture model to the problem of discovering new species.
- Treating word labels for categories as “just another feature”.

Initializing new concepts

- When is a new observed instance different enough from our previous experience to posit a new concept, species, or kind?
- A ubiquitous problem in cognitive development, everyday learning, and scientific discovery.
 - The “new species” problem
 - The “none of the above” problem
 - A version of Fodor’s problem?

Anderson's solution

- Likelihood $p(x|c)$ for new concept: high.
 - Parameters can be tuned to fit x perfectly.
- Prior $p(c)$ for new concept: low
 - Dirichlet (“rich get richer”) process.

- For old concept j (with n_j previous instances):

$$p(c_j) = \frac{n_j}{n + \gamma} \quad n = \sum_j n_j$$

- For new concepts:

$$p(\text{new concept}) = \frac{\gamma}{n + \gamma}$$

- Prior probability that a new concept exists decreases as more things are seen.

Two interesting issues

- Applying an infinite mixture model to the problem of discovering new species.
- Treating word labels for categories as “just another feature”.

Advantages of “labels as features”

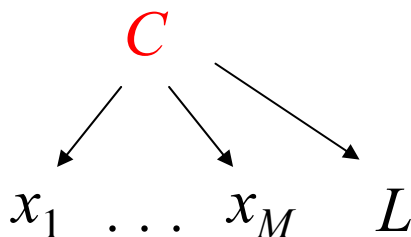
- Parsimony / Continuity
 - Learning words is just like any other kind of feature learning
- Flexibility
 - Can learn multimodal or nonlinearly separable lexical classifications, by associating the same label with multiple concepts.
- What do you think?

Problems with “labels as features”

- Words are very strong features
 - “is a mammal” versus “breathes air”
- Words and features pattern very differently across categories:
 - Features cluster: the more features are shared by a category’s members, the more likely new features are to be shared.
 - Words “anti-cluster”: map ~1-to-1 onto concepts, so learning a label for a concept makes it *less likely* that new labels will refer to that concept.
- Words are conventions, created tools for communication and reference.

Fixing the problem?

- Words as an invitation to form categories.
- Category labels are rigid designators.
- Modified Anderson
 - Infinite number of states for L .
 - One-to-one deterministic mapping between states of C and states of L .



Inductive reasoning in biology

- A historical shift in cognitive research
 - From inductive learning of simple abstract categories (1950's-1985ish) to inductive generalization with complex real-world knowledge structures (1985ish-present).
- The computational challenge
 - Bring statistical learning models together with structured knowledge representations

Which argument is stronger?

Cows have biotinic acid in their blood

Horses have biotinic acid in their blood

Rhinos have biotinic acid in their blood

All mammals have biotinic acid in their blood

Cows have biotinic acid in their blood

Dolphins have biotinic acid in their blood

Squirrels have biotinic acid in their blood

All mammals have biotinic acid in their blood

“Diversity phenomenon”

Bayes plus theories

- Rational statistical inference (Bayes):

$$p(h | d) = \frac{p(d | h) p(h)}{\sum_{h' \in H} p(d | h') p(h')}$$

- Domain theories generate the necessary ingredients: hypothesis space H , priors $p(h)$.
 - Well-matched to the structure of the natural world.
 - Learnable from limited data.
 - Computationally tractable inference.

Alternative models

- Bayesian models
 - With different generative theories
 - Without generative theories
- Alternative models
 - Similarity-based
 - Pure theory-based

The computational problem

Species 1	●
Species 2	?
Species 3	?
Species 4	?
Species 5	?
Species 6	?
Species 7	?
Species 8	?
Species 9	●
Species 10	?

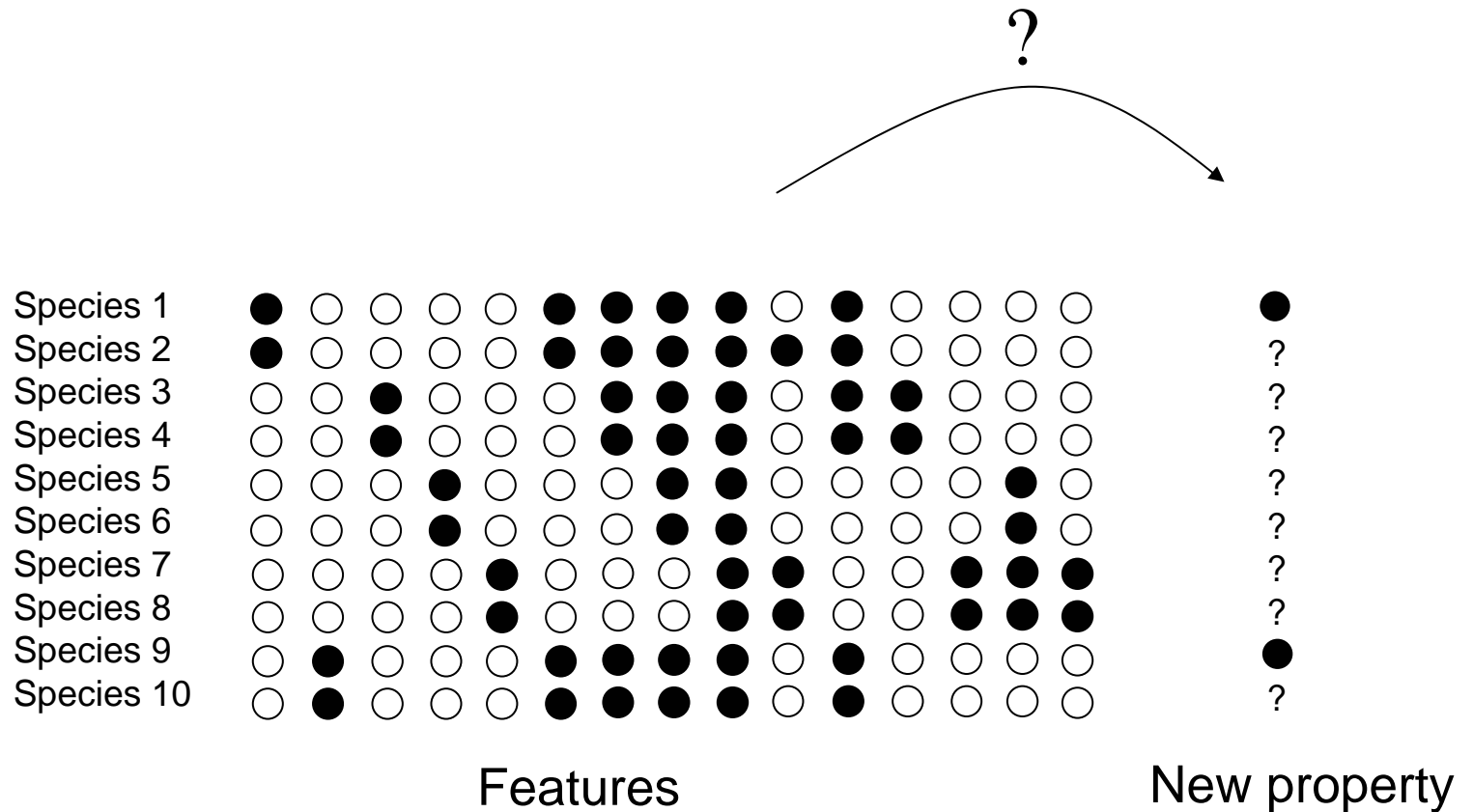
New property

Feature rating data

Osherson, D. N., et. al. "Category-based Induction." *Psychological Review* 197 (1990): 185-200.

- People were given 48 animals, 85 features, and asked to rate whether each animal had each feature.
- E.g., elephant:
 - 'gray' 'hairless' 'toughskin'
 - 'big' 'bulbous' 'longleg'
 - 'tail' 'chewteeth' 'tusks'
 - 'smelly' 'walks' 'slow'
 - 'strong' 'muscle' 'quadrapedal'
 - 'inactive' 'vegetation' 'grazer'
 - 'oldworld' 'bush' 'jungle'
 - 'ground' 'timid' 'smart'
 - 'group'

The computational problem



Similarity-based models

(Osherson et al.)

- 20 subjects rated the strength of 45 arguments:

X_1 have property P.

X_2 have property P.

X_3 have property P.

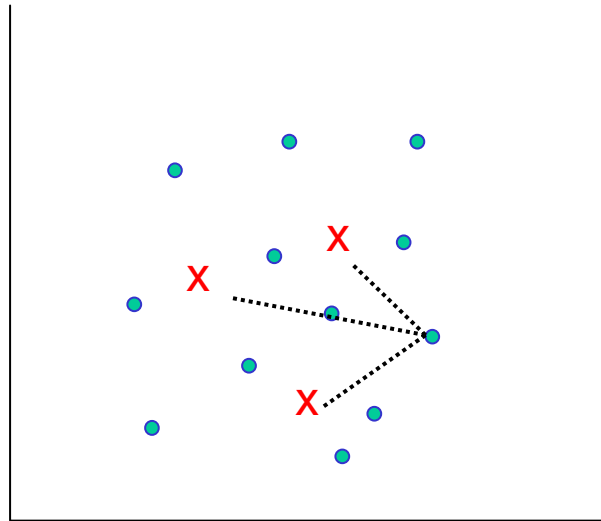
All mammals have property P.

- 40 different subjects rated the similarity of all pairs of 10 mammals.

Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$



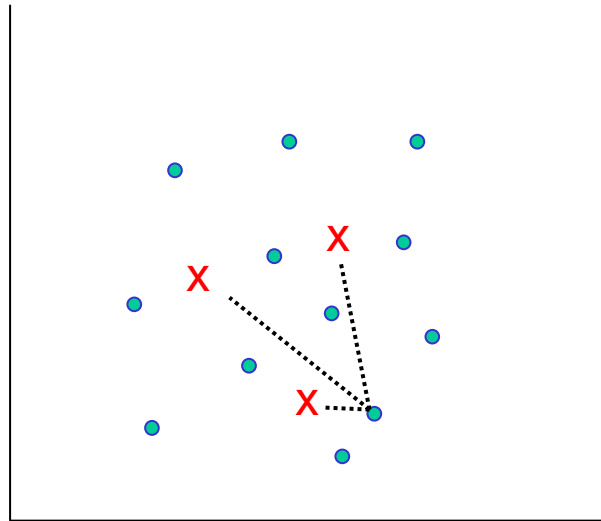
Mammals: •

Examples: x

Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$



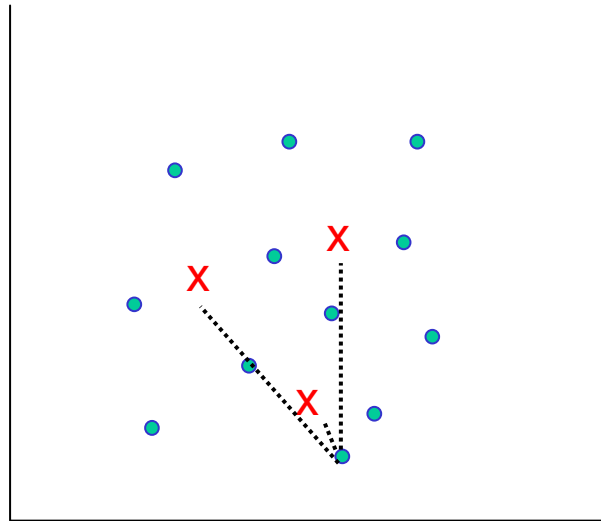
Mammals: •

Examples: x

Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$



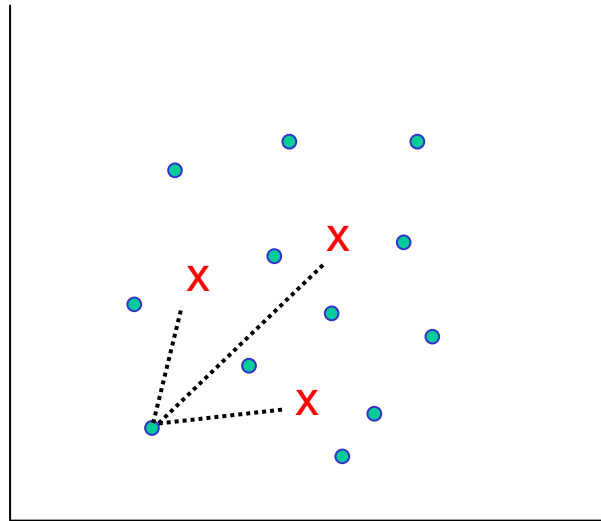
Mammals: •

Examples: x

Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$



Mammals: •

Examples: x

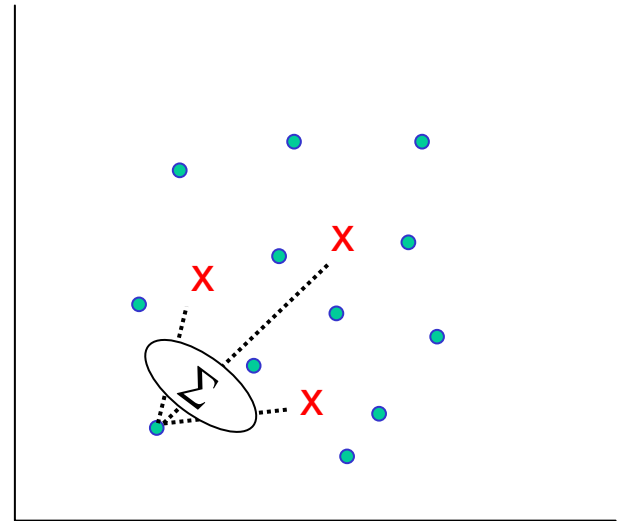
Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$

- **Sum-Similarity:**

$$\text{sim}(i, X) = \sum_{j \in X} \text{sim}(i, j)$$



Mammals: •

Examples: x

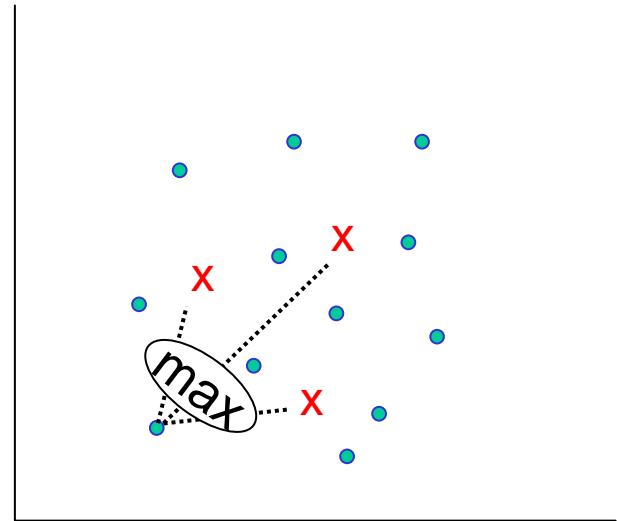
Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$

- **Max-Similarity:**

$$\text{sim}(i, X) = \max_{j \in X} \text{sim}(i, j)$$



Mammals: •

Examples: x

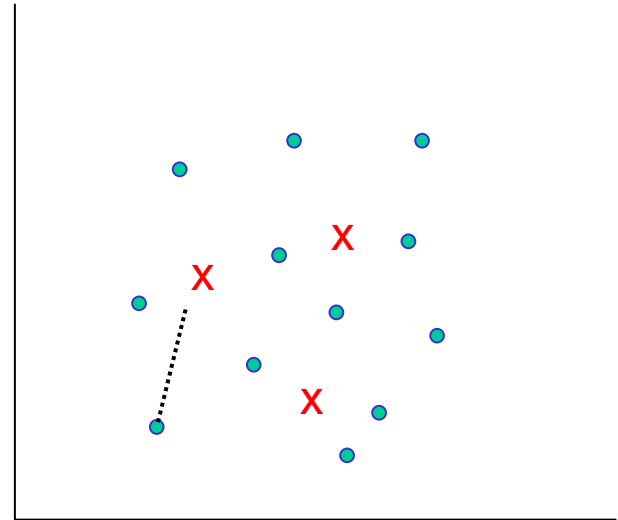
Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$

- **Max-Similarity:**

$$\text{sim}(i, X) = \max_{j \in X} \text{sim}(i, j)$$



Mammals: •

Examples: x

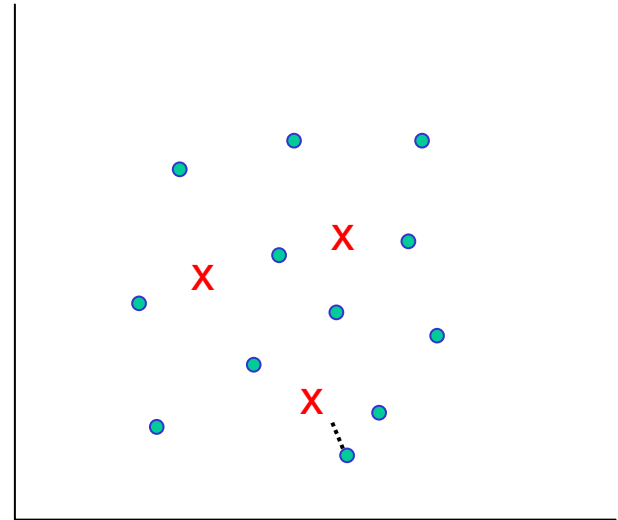
Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$

- **Max-Similarity:**

$$\text{sim}(i, X) = \max_{j \in X} \text{sim}(i, j)$$



Mammals: •

Examples: x

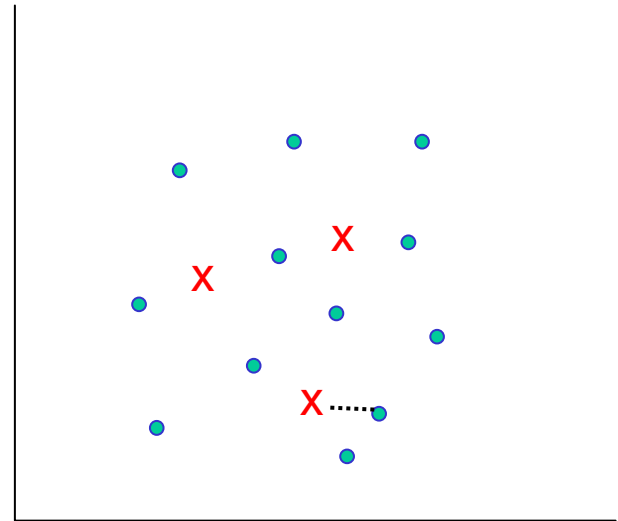
Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$

- **Max-Similarity:**

$$\text{sim}(i, X) = \max_{j \in X} \text{sim}(i, j)$$



Mammals: •

Examples: x

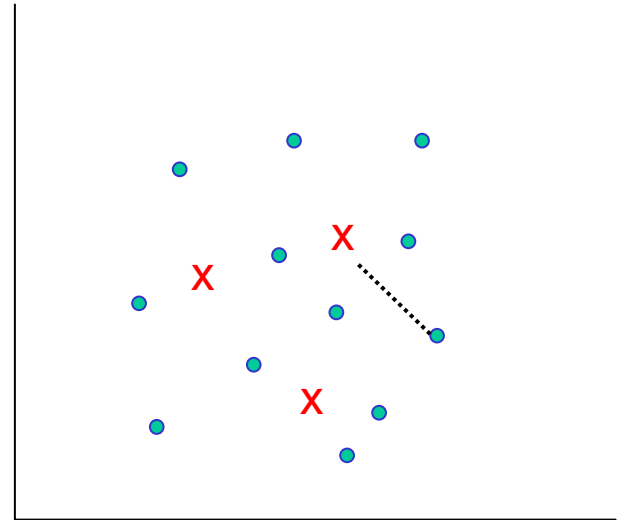
Similarity-based models

(Osherson et al.)

$$\text{strength}(\text{“all mammals”} \mid X) = \sum_{i \in \text{mammals}} \text{sim}(i, X)$$

- **Max-Similarity:**

$$\text{sim}(i, X) = \max_{j \in X} \text{sim}(i, j)$$



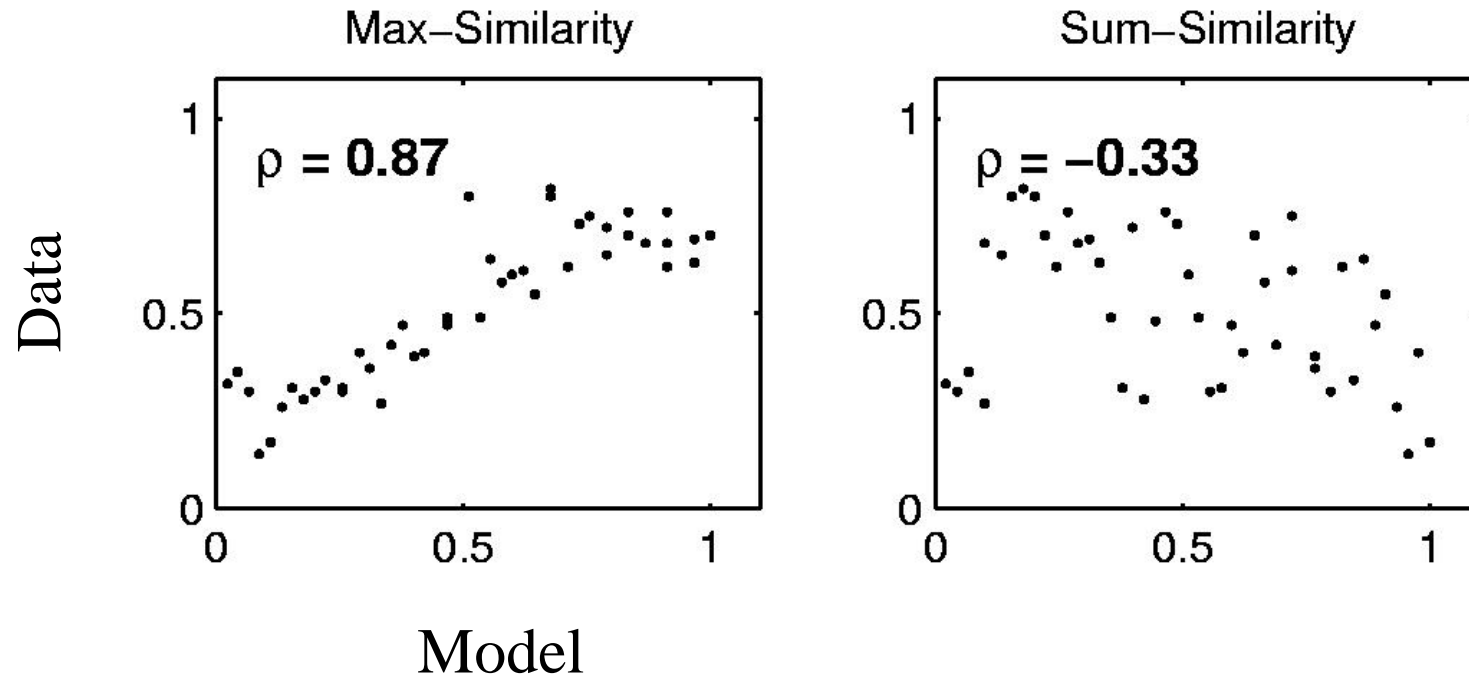
Mammals: •

Examples: x

Sum-sim versus Max-sim

- Two models appear functionally similar:
 - Both increase monotonically as new examples are observed.
- Reasons to prefer Sum-sim:
 - Standard form of exemplar models of categorization, memory, and object recognition.
 - Analogous to kernel density estimation techniques in statistical pattern recognition.
- Reasons to prefer Max-sim:
 - Fit to generalization judgments

Data vs. models



Each “.” represents one argument:

- X_1 have property P.
- X_2 have property P.
- X_3 have property P.

All mammals have property P.

Three data sets

Images removed due to
copyright considerations.

Conclusion

kind: “all mammals”

“horses”

“horses”

Number of
examples:

3

2

1, 2, or 3

Feature rating data

(Osherson and Wilkie)

- People were given 48 animals, 85 features, and asked to rate whether each animal had each feature.
- E.g., elephant:
 - 'gray' 'hairless' 'toughskin'
 - 'big' 'bulbous' 'longleg'
 - 'tail' 'chewteeth' 'tusks'
 - 'smelly' 'walks' 'slow'
 - 'strong' 'muscle' 'quadrapedal'
 - 'inactive' 'vegetation' 'grazer'
 - 'oldworld' 'bush' 'jungle'
 - 'ground' 'timid' 'smart'
 - 'group'

Three data sets

Max-sim

Images removed due to
copyright considerations.

Sum-sim

Conclusion

kind:	“all mammals”	“horses”	“horses”
-------	---------------	----------	----------

Number of
examples:

3

2

1, 2, or 3

Problems for sim-based models

- No principled explanation for why similarity to examples should follow Max-Sim.
 - Leading similarity-based models of categorization, memory, and object recognition use Sum-Sim.
- Free parameters mixing similarity and coverage terms, and possibly Max-Sim and Sum-Sim terms.
- Does not extend to induction with other kinds of properties, e.g., from Smith et al., 1993:

Dobermanns can bite through wire.

German shepherds can bite through wire.

Poodles can bite through wire.

German shepherds can bite through wire.

Marr's Three Levels of Analysis

- Computation:
 - “What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?”
- Representation and algorithm:
 - Max-sim, Sum-sim
- Implementation:
 - Neurobiology

Theory-based induction

- Scientific biology: species generated by an evolutionary branching process.
 - A tree-structured taxonomy of species.

Images removed due to
copyright considerations.

- Taxonomy also central in folkbiology (Atran).

Theory-based induction

Begin by reconstructing intuitive taxonomy
from similarity judgments:

Images removed due to
copyright considerations.

How taxonomy constrains induction

- Atran (1998): “Fundamental principle of systematic induction” (Warburton 1967, Bock 1973)
 - Given a property found among members of any two species, the best initial hypothesis is that the property is also present among all species that are included in the smallest higher-order taxon containing the original pair of species.

Image removed due to
copyright considerations.

Cows have property P.
Dolphins have property P.
Squirrels have property P.

All mammals have property P.

Strong (0.76 [max = 0.82])

Image removed due to
copyright considerations.

Cows have property P.
Dolphins have property P.
Squirrels have property P.

All mammals have property P.

Strong: 0.76 [max = 0.82]

Cows have property P.
Horses have property P.
Rhinos have property P.

All mammals have property P.

Weak: 0.17 [min = 0.14]

Image removed due to
copyright considerations.

Cows have property P.
Dolphins have property P.
Squirrels have property P.

All mammals have property P.

Strong: 0.76 [max = 0.82]

Seals have property P.
Dolphins have property P.
Squirrels have property P.

All mammals have property P.

Weak: 0.30 [min = 0.14]

Taxonomic
distance

Max-sim

Images removed due to
copyright considerations.

Sum-sim

Conclusion

kind: “all mammals”

“horses”

“horses”

Number of
examples:

3

2

1, 2, or 3

The challenge

- Can we build models with the best of both traditional approaches?
 - Quantitatively accurate predictions.
 - Strong rational basis.