Subsumption

Motivation

Imagine you're organizing a library. Traditional point embeddings are like placing books at specific coordinates—you can measure how "close" two books are, but you can't naturally represent that "Mystery Novels" **contains** "Agatha Christie Novels" which **contains** "Hercule Poirot Stories".

Box embeddings solve this elegantly: represent each category as a **box** (a region in space). If the "Mystery Novels" box completely contains the "Agatha Christie" box, then we've encoded the containment relationship geometrically. This is the central idea: logical subsumption becomes geometric containment.

The "Russian doll" intuition: Just as Russian nesting dolls fit inside each other, boxes can nest. If box A contains box B, and box B contains box C, then box A must contain box C—this transitivity is built into the geometry itself. No need for explicit rules or constraints; the mathematics enforces consistency automatically.

This geometric interpretation of logical subsumption is powerful because:
- It provides an intuitive visual representation of hierarchical relationships (you can literally "see" the hierarchy)
- Containment can be computed efficiently using geometric operations (just check if boundaries satisfy inequalities)
- The probabilistic extension to Gumbel boxes maintains this intuitive structure while enabling gradient-based learning (the boxes can "wiggle" slightly, making optimization possible)

Historical context: Box embeddings emerged in 2018 (Vilnis et al.) as a solution to a fundamental limitation of point-based embeddings: points cannot naturally represent containment relationships. Traditional vector embeddings (Word2Vec, GloVe, TransE) represent entities as points in space, where distance measures similarity but cannot encode hierarchy. Box embeddings solve this by representing entities as regions (boxes), where containment becomes geometric. This shift from points to regions enables natural representation of hierarchical structures like taxonomies, ontologies, and knowledge graphs.

Why geometric containment? The choice of geometric containment as a representation for logical subsumption is not arbitrary—it's a natural encoding that has deep connections to set theory and category theory. In set theory, subsumption (subset relationship) is the fundamental ordering relation, and geometric containment in Euclidean space provides a continuous, learnable analog. The key insight is that containment is **transitive**: if A contains B and B contains C, then A contains C. This transitivity property is essential for representing hierarchical structures like taxonomies, where "animal" contains "mammal" which contains "dog". Box embeddings preserve this transitivity through geometric containment, ensuring that learned hierarchies are consistent and interpretable.

Connection to category theory: In category theory, subsumption corresponds to a morphism between objects, where the containment relationship defines a partial order. Box embeddings provide a concrete geometric realization of this abstract structure, making it computable and learnable. The geometric interpretation also connects to the concept of **topological spaces**, where containment defines neighborhoods and continuity, providing a bridge between discrete logical structures and continuous geometric representations.

Definition

Subsumption is a fundamental concept in formal logic. In box embeddings, we define subsumption as a binary relation: box A **subsumes** box B if and only if B is geometrically contained within A, denoted $B \subseteq A$.

For hard boxes, subsumption is deterministic: either $B \subseteq A$ (subsumption holds with probability 1) or not (probability 0). For Gumbel boxes, subsumption becomes probabilistic, measured by the containment probability $P(B \subseteq A)$ (see the Containment Probability document). We say A subsumes B when $P(B \subseteq A) \approx 1$ (typically > 0.99 in practice).

Statement

> **Theorem (Geometric Subsumption).** For boxes A and B, the following are equivalent:
>
> 1. Box A subsumes box B
> 2. $B \subseteq A$ (geometric containment)
> 3. $P(B \subseteq A) = 1$ for hard boxes, or $P(B \subseteq A) \approx 1$ for Gumbel boxes
>
> Geometric containment directly corresponds to logical subsumption.

Proof

We prove that geometric containment implies logical subsumption. Under the uniform base measure on $[0, 1]^d$, the containment probability is defined as:

$$P(B \subseteq A) = \frac{\text{Vol}(A \cap B)}{\text{Vol}(B)}$$

This measures the fraction of box B's volume that lies within box A.

Hard boxes: When $B \subseteq A$ (geometric containment), we have $A \cap B = B$ by definition of intersection. Therefore $\text{Vol}(A \cap B) = \text{Vol}(B)$, which gives $P(B \subseteq A) = \frac{\text{Vol}(B)}{\text{Vol}(B)} = 1$. This establishes that geometric containment implies subsumption with probability 1.

Conversely, if $P(B \subseteq A) = 1$, then $\text{Vol}(A \cap B) = \text{Vol}(B)$, which means every point in B is also in A, so $B \subseteq A$. This completes the equivalence for hard boxes.

Gumbel boxes: For Gumbel boxes with random boundaries, we use the first-order Taylor approximation (see the Containment Probability document):

$$P(B \subseteq A) \approx \frac{E[\text{Vol}(A \cap B)]}{E[\text{Vol}(B)]}$$

When the expected boundaries satisfy $E[B] \subseteq E[A]$ (where $E[B]$ denotes the box with expected boundaries), and when β is small, the approximation gives $P(B \subseteq A) \approx 1$, establishing probabilistic subsumption.

Interpretation

Subsumption provides a unified framework for encoding three fundamental logical and semantic relationships:

1. **Entailment**: If premise box P subsumes hypothesis box H, then P entails H. The geometric containment directly captures the logical relationship that all instances satisfying the premise also satisfy the hypothesis.

2. **Concept hierarchies**: Parent concepts subsume child concepts. For example, "animal" subsumes "dog" because every dog is an animal. The parent box (animal) must contain the child box (dog) in the embedding space.

3. **Logical consequence**: The containment relationship represents logical subsumption in a continuous, learnable form. Unlike discrete logical systems, box embeddings allow for soft, probabilistic subsumption that can be optimized through gradient descent.

Partial subsumption: When $0 < P(B \subseteq A) < 1$, we have **partial subsumption**. This occurs when boxes overlap but B is not fully contained in A. The probability value quantifies the degree of containment: $P(B \subseteq A) = 0.8$ means 80% of box B's volume lies within box A. This soft subsumption enables modeling of uncertain or graded logical relationships, which is particularly useful for learning from noisy or ambiguous data.
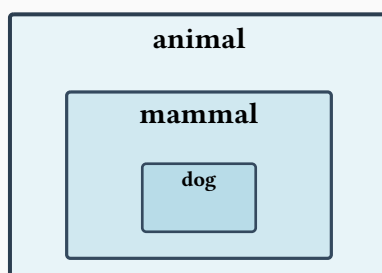
Example

A concrete puzzle: Can you arrange three boxes—"dog", "mammal", and "animal"—so that "dog" is inside "mammal", and "mammal" is inside "animal"?

The answer is yes, and here's how box embeddings do it:

Consider the hierarchy: "dog" $\subseteq$ "mammal" $\subseteq$ "animal"

Visual representation: The nested box structure is shown below, where each box represents a concept and containment encodes subsumption. The diagram illustrates how geometric containment directly encodes the logical hierarchy.

Box embeddings (hard boxes):
- "dog" = box from $[0.2, 0.4]$ to $[0.4, 0.6]$ (volume = 0.04) — a small, precise box
- "mammal" = box from $[0.1, 0.3]$ to $[0.5, 0.7]$ (volume = 0.16) — a larger box that completely contains "dog"
- "animal" = box from $[0.0, 0.0]$ to $[1.0, 1.0]$ (volume = 1.0) — the entire space, containing everything

Verification (the "containment test"):
- Intersection of "dog" and "mammal": $[0.2, 0.4]$ to $[0.4, 0.6]$ = "dog" (the entire "dog" box)
- $P(\text{dog} \subseteq \text{mammal}) = \frac{\text{Vol(dog} \cap \text{mammal)}}{\text{Vol(dog)}} = \frac{0.04}{0.04} = 1.0$ ✓
- Intersection of "mammal" and "animal": $[0.1, 0.3]$ to $[0.5, 0.7]$ = "mammal" (the entire "mammal" box)
- $P(\text{mammal} \subseteq \text{animal}) = \frac{\text{Vol(mammal} \cap \text{animal)}}{\text{Vol(mammal)}} = \frac{0.16}{0.16} = 1.0$ ✓

The "aha!" moment: Notice that containment is geometrically explicit—you can literally draw these boxes and see the hierarchy. The parent box (animal) contains the child box (mammal),

which in turn contains its child (dog), creating a clear hierarchical structure in the embedding space. This is not just a metaphor; it's the actual mathematical structure.

Notes

Beyond hard boxes: The deterministic containment shown above extends naturally to probabilistic settings. Gumbel boxes allow partial containment, where $0 < P(B \subseteq A) < 1$, enabling modeling of uncertain or graded relationships. This probabilistic extension preserves the geometric intuition while enabling gradient-based learning.

Connection to knowledge graphs: Box embeddings have been successfully applied to WordNet (82,114 entities, 84,363 edges), achieving F1 scores above 90% for hypernym prediction. The geometric containment directly encodes the IS-A relationship, making box embeddings particularly well-suited for hierarchical knowledge representation.

Future directions: Current work explores extending box embeddings to more complex geometric structures (balls, cones, arbitrary regions) and investigating connections to hyperbolic geometry. The fundamental insight—that containment relationships are naturally geometric—suggests many unexplored applications in knowledge representation and reasoning.