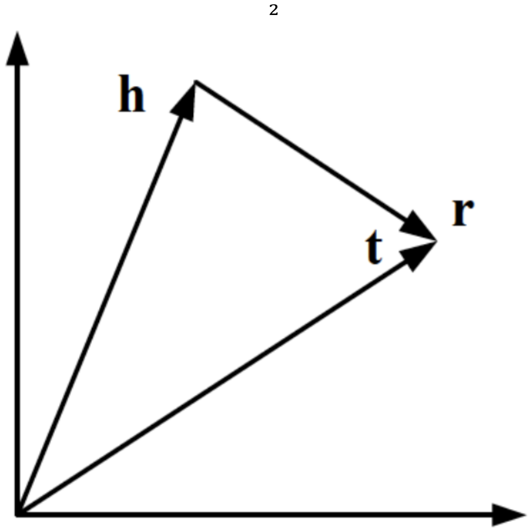
HOMEWORK 04: Link Prediction & Graph Embedding

StudentID: 20127054

Name: Ngô Văn Trung Nguyên

a) ***Learn and present the TransE model the way you understand it.***
   - The TransE model is a translational distance model that represents entities and relations as vectors in a shared semantic space of reduced dimension R^d, where d refers to the target space dimension.
   - In this model, a fact in the source space is represented as a triplet (h, r, t) where h stands for head, r stands for relation and t stands for tail. The model interprets the relationship between entities as a translation vector, ensuring that connected entities in the embedding space have a small distance between them.
   - In terms of vector computation, it implies that the addition of head and relation vector should be approximate to the tail vector, written as $h + r \approx t$.
   - Example:
     - $h_1 = emb("Ottawa"), t_1 = emb("Canada")$
     - $h_2 = emb("Berlin"), t_1 = emb("Germany")$
     - $r = "CapilatOf"$
     - From rows above, we have $h_1 + r \approx t_1$ and $h_2 + r \approx t_2$.
   - TransE performs linear transformation and the scoring function is negative distance between $h + r$ and t, or $f = -\|h + r - t\|_{\frac{1}{2}}$



b) *Suppose we consider a simpler loss function for the TransE model as follows:*

$$L_{simple} = \sum_{(h,r,t)\in S} d(h + r, t)$$

Whether the entity and relational embeddings become better than the original version after minimizing the loss function to zero. Give an example to support that statement.

- No, using the loss function L_{simple} would likely not lead to better embeddings than the original TransE.
- The main drawback of L_{simple} is that it does not distinguish between correct and incorrect triplets during training. It simply tries to minimize the distance between embeddings for all triplets in the training set, regardless of whether they are valid knowledge graph facts or not.
- This could cause the model to learn embeddings that do not properly represent semantic relations.
- Example:
  - Consider these correct facts:
    - (Tokyo, locatedIn, Japan)
    - (Paris, locatedIn, France)
  - Incorrect facts:
    - (Tokyo, locatedIn, France)
  - Minimizing L_{simple}would push the embeddings of (Tokyo, locatedIn, France) close to each other, even though it is not a true statement.
  - In contrast, TransE's original loss function incorporates the difference between correct and incorrect triplets. This guides the model to properly embed entities and relations according to their semantic meaning.
- In conclusion, L_{simple} is too simplistic as it could cause the model to learn representations that do not properly reflect the true relations in the knowledge graph., unlike TransE's original function which considers both correct and incorrect facts.

c) *Suppose we consider a simpler loss function for the TransE model as follows:*

$$L_{no\ margin} = \sum_{(h,r,t)\in S} \sum_{(h',r',t')\notin S} max[0, d(h + r, t) - d(h' + r, t')]$$

Whether the entity and relational embeddings become better than the original version after minimizing the loss function to zero. Give an example to support that statement.

- No, using the proposed simpler loss function $L_{no\ margin}$ would not necessarily lead to better entity and relational embeddings compared to the original TransE model.
- The key difference between $L_{no\ margin}$ and the original TransE loss function is the removal of the margin term.
- Without a margin, the loss function only penalizes incorrect triplet predictions that have a higher distance than the correct triplet, but provides no incentive for the correct triplet to be clustered together and separated from incorrect ones by a margin.

This could lead to embeddings that minimize the loss by spreading out all predictions rather than tightly clustering correct triplets.

- Example: consider A, B, C and r where (A,r,B) is correct but (A,r,C) is incorrect.
  - With just $L_{no\ margin}$, the model could learn embeddings where:
    - d(A+R, B) = 1
    - d(A+R, C) = 2
  - This minimizes the loss, but the correct and incorrect triplets are not well separated.
  - With the margin in the original TransE, the model would be incentivized to learn:
    - d(A+R, B) = 1
    - d(A+R, C) > 1 + γ
- Therefore, simply minimizing $L_{no\ margin}$ to zero does not guarantee better embeddings compared to the original TransE formulation. The margin plays an important role in learning high quality representations.

d) Do entities embedding in TransE need to be normalized to the same length? Why ?

- Entities embedding in TransE need to be normalized to the same length so that entities are not pushed to have large embeddings, which would result in the model having difficulty in distinguishing positive and negative triplets.

e) Give an example of a simple graph for which no perfect embedding exists, i.e., no embedding perfectly satisfies $\boldsymbol{h} + \boldsymbol{r} = \boldsymbol{t}$ for all $(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}) \in \boldsymbol{S}$ and $\boldsymbol{h}' + \boldsymbol{r} \neq \boldsymbol{t}'$ for $(\boldsymbol{h}', \boldsymbol{r}, \boldsymbol{t}') \notin \boldsymbol{S}$, for any choice of entity embeddings (e for e ∈ E) and relationship embeddings (r for r ∈ R). Explain why this graph has no perfect embedding in this system.

- The graph consists of 4 nodes: A, B, C, D
- It contains 2 reltions: r1, r2
- It has the following $S = \{(A, r1, B), (\mathrm{C}, r2, \mathrm{D}), (\mathrm{A}, \mathrm{r1}, \mathrm{D})\}$
- This graph does not have a perfect embedding because there is no way to assign embeddings to the entities A, B, C, D and relationships r1, r2 such that:
  - $h + r = t$ for the facts in S.
  - $h' + r \neq t'$ for any other facts not in S.
- $(\mathrm{A}, \mathrm{r1}, \mathrm{B})$ is in S, so $A + r1 = \mathrm{B}$
- $(\mathrm{C}, r2, \mathrm{D})$ is in S, so $\mathrm{C} + r2 = D$
- $(\mathrm{A}, \mathrm{r1}, \mathrm{D})$ is in S, so $\mathrm{A} + r1 = D$
- .However $(\mathrm{A}, \mathrm{r1}, \mathrm{C})$ is not in S, so $\mathrm{A} + r1 \neq \mathrm{C}$