Density-Based Document Ranking

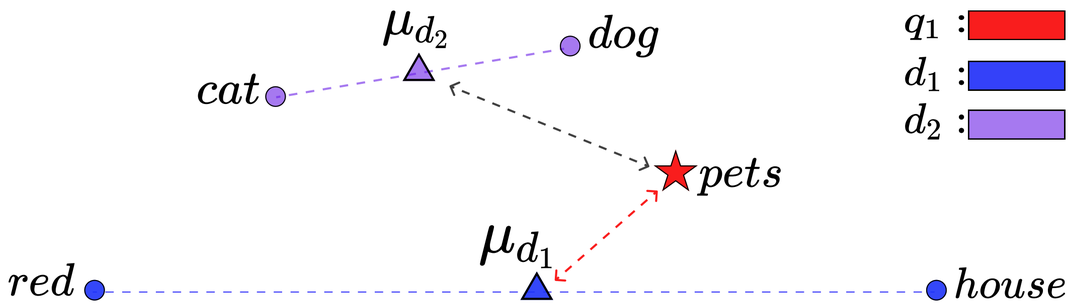
Thesis Proposal

giovanni.stilo@univaq.it, andrea.dangelo6@graduate.univaq.it

Document Ranking

Problem statement

- Given a query Q and a corpus of documents D, rank the documents based on the relevance with the query.
- Document Ranking often use LLM-based rankers for this task. They are usually fine-tuned on the Mean of embeddings between query and document.



Density-based retrieval

What we found

- We found that rewarding dense subareas in the latent space is more effective for Document Ranking than the Mean of Embeddings.
- Thus, can we fine-tune an LLM with a similar loss function to improve Ranking effectiveness?

$$DbU(C_Q, C_D, k) = \sum_{\mathbf{e_q} \in C_Q} \sum_{\mathbf{e_d} \in \mathcal{A}(\mathbf{e_q}, C_D, k)} \frac{\text{local-density}(\mathbf{e_q}, \mathbf{e_d}, C_D, k)}{|C_Q| \cdot |\mathcal{A}(\mathbf{e_q}, C_D, k)|}$$