

Dynamic Interpreter Structure

Inputs :

- base model weights, biases, gradients
- dataset feature name embeddings

Output :

- cosine similarities between each feature embedding and target embedding

