Author - Manav Mittal (2021538)

Question 1 Report

LLAMA 3.1

Self Inconsistency

- LLAMA was inconsistent on Mathematical and Reasoning questions.
- Just by Paraphrasing the prompt we can see differnet results which should not happen since the meaning of the prompt is same

Fact Checking

- LLAMA is failing on General Knowledge questions (not obvious ones but some very nit-picked questions)

OpenHathi

Self Inconsistency

- OpenHathi failed on basic arithmatic operations like square, square-root and multiplication
- It is unable to perform reasoning and giving different answers if the prompt is paraphrased

Fact Checking

- It is giving wrong answers for general knowledge questions

Question 2 Report

Evaluation and Probing Results

Linear Regression Head

First Layer Embeddings

- Training MSE - 4.4467268811967e-25
- Testing MSE - 4778.288728600571

Middle Layer Embeddings

- Training MSE - 1.0223085129030658e-26
- Testing MSE - 324.91917257035783

Last Layer Embeddings

- Training MSE - 1.085693486637574e-26

- Testing MSE - 230.72164799163102

Analysis

- We can clearly see that as we take embeddings of deeper layers of the model the performace improves for both Training and Testing data. This clearly shows that LLM has learned well and performance improves as we go deep in the network.
- As we go deep in the network we can see the MSE (Mean Square Error) loss decreasing.
- Model is able to predict Popularity score of the song more precisely on deeper layers of the LLM.

Classification Head

First Layer Embeddings

- Training Accuracy - 0.98125
- Testing Accuracy - 0.625

Middle Layer Embeddings

- Training Accuracy - 1.0
- Testing Accuracy - 0.925

Last Layer Embeddings

- Training Accuracy - 1.0
- Testing Accuracy - 0.975

Analysis

- This is same as what we saw in Analysis of Linear Regression Head
- We can clearly see that as we take embeddings of deeper layers of the model the performace improves for both Training and Testing data This clearly shows that LLM has learned well and performance improves as we go deep in the network.
- As we go deep in the network we can see the classification accuracy increasing.
- Model is able to predict genre of the song more precisely on deeper layers of the LLM.

Discussion

Findings

- LLM is performing very well in predicting the popularity score and genre of the song.
- It is only provided with the Track name, the album name to which the song belongs and the artist name. And only with these three feilds it is predicting popularity score and genre of the exceptionally well.

Patterns

- As we go deeper in the network the quality of the embeddings we are getting is improving. Which implies that at each and every layer the model is learning something.