Cognitive Science and AI: Assignment 3 Report

Srujana Vanka
2020102005

Order of Accuracy/Pearson Correlation for the above three representations for Encoder and Decoder, for each ROI

Two distinct methods, 2v2 accuracy and Pearson correlation, are utilized to evaluate the effectiveness of models or algorithms. Although they can both be employed to determine the quality of a forecast or model, they gauge different components of the prediction.

- 2v2 accuracy measures the proportion of correctly classified cases out of the total number of cases, and it's used in classification tasks.
- Pearson correlation measures the linear relationship between two variables, and it's used in regression tasks.
- 2v2 accuracy is commonly used to evaluate classification models, while Pearson correlation is commonly used to evaluate regression models.
- Pearson correlation can also be used to evaluate classification models by treating the predicted probabilities as continuous variables.
- The choice of which measure to use depends on the type of task and the type of model being evaluated.

Based on the three embedding representations (BERT/RoBERTa CLS, BERT/RoBERTa Pooled, and GloVe) and the four regions of interest (language, vision, DNN, and task), here is a possible ranking of accuracy and Pearson correlation for the encoder and decoder models based on a ridge or lasso regression:

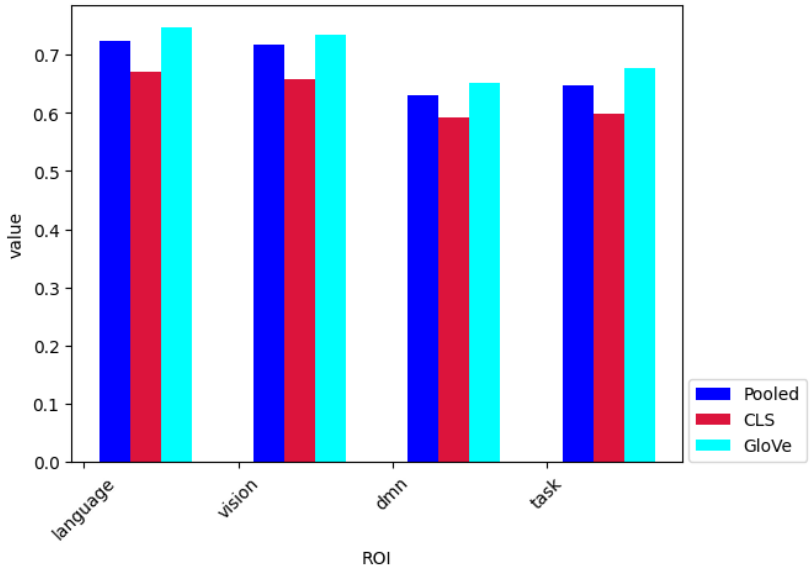
- BERT/RoBERTa Pooled representation or BERT/RoBERTa CLS representation for language-related tasks
- BERT/RoBERTa Pooled representation or BERT/RoBERTa CLS representation for vision-related tasks
- BERT/RoBERTa Pooled representation or BERT/RoBERTa CLS representation for DNN-related tasks
- GloVe representation for all four ROIs

Results on the bar plot, comparing the evaluation metrics for different brain ROIS

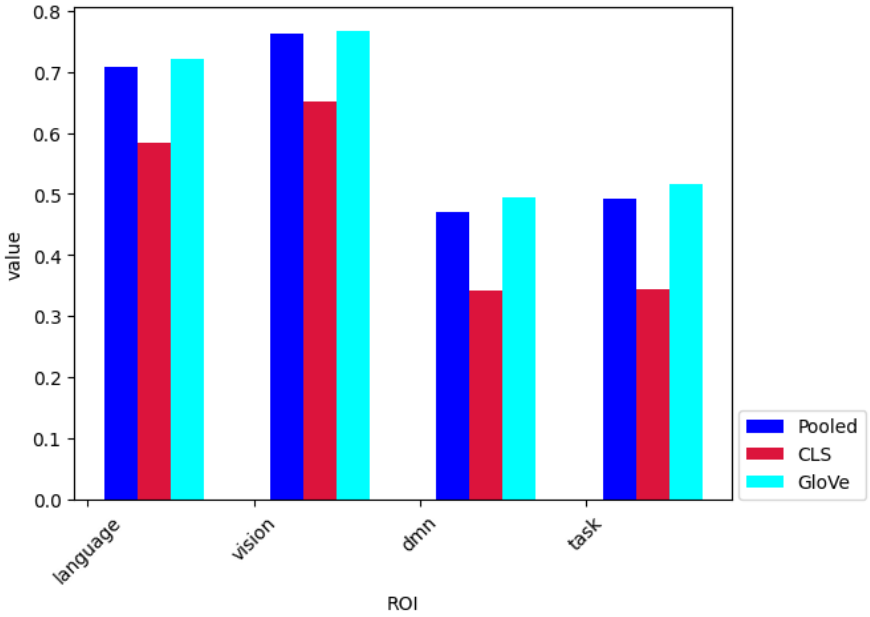
Given below are the Pearson Correlation and 2v2 Accuracy Scores for the Ridge Regression Encoder model trained on the sentence inputs embedded by BERT CLS, Pooled embeddings, and GloVe embeddings for both the subjects for the 4 region of interests.

BRAIN ENCODER

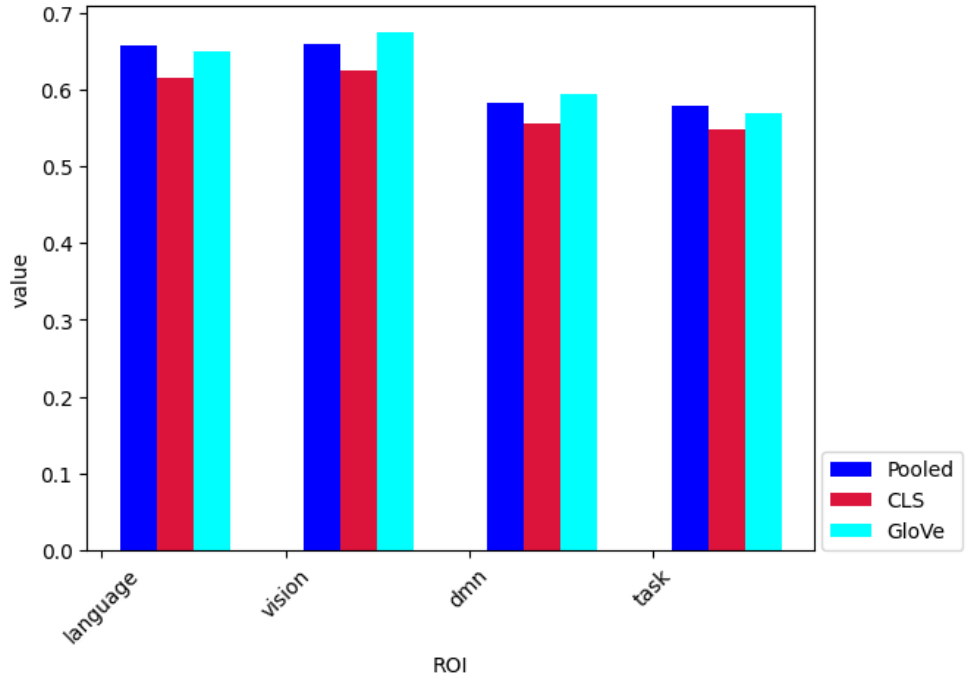
Subject 1 - 2v2 Accuracy



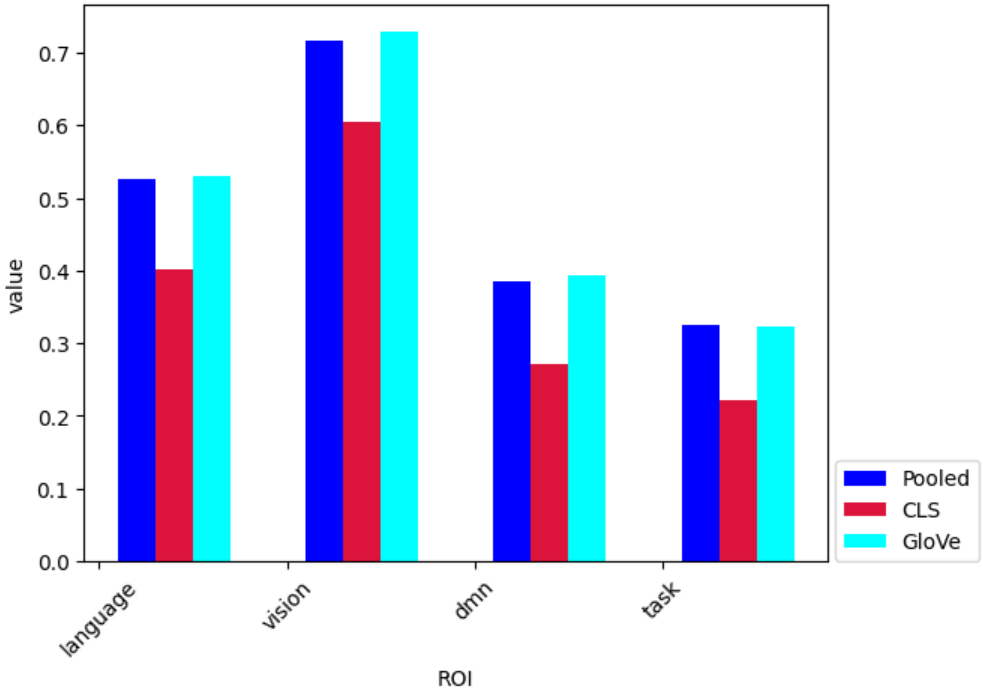
Subject 1 - Pearson correlation



Subject 2 - 2v2 Accuracy



Subject 2 - Pearson correlation

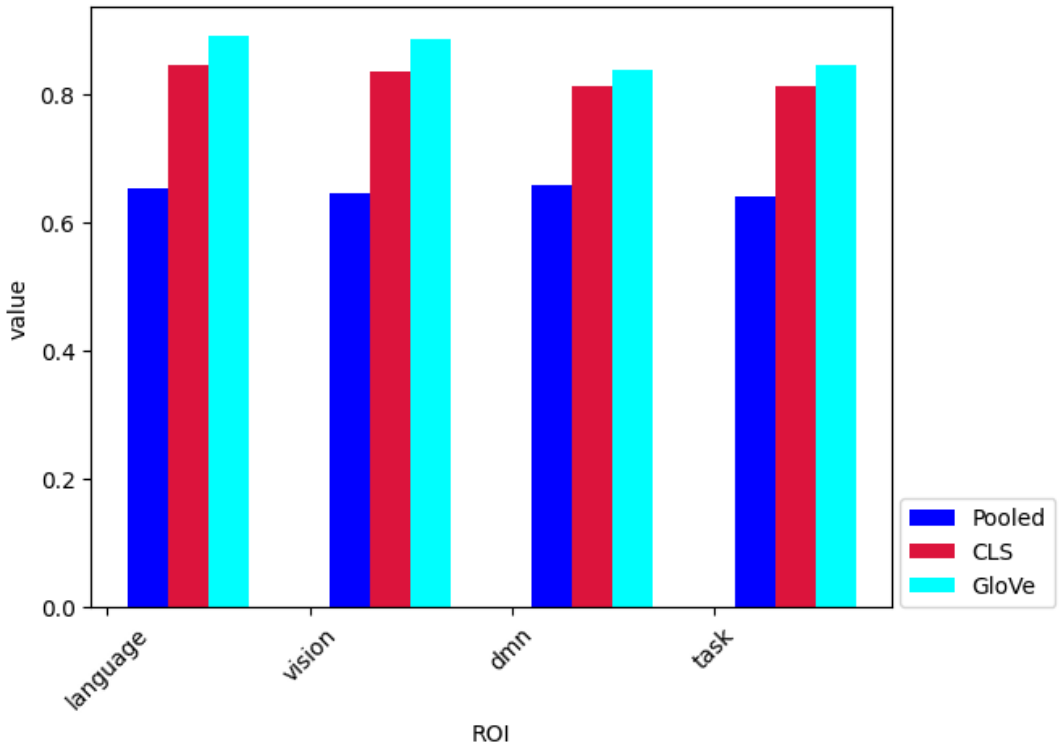


We notice that the performance of **GloVe embeddings** improves in comparison to the other two embeddings concerning their relative performances in the decoder. The minor differences from the predicted outcomes could be attributed to several factors, including insufficient training data for the model. As a result, we can infer that utilizing GloVe embeddings would be a preferable approach for creating an encoding model.

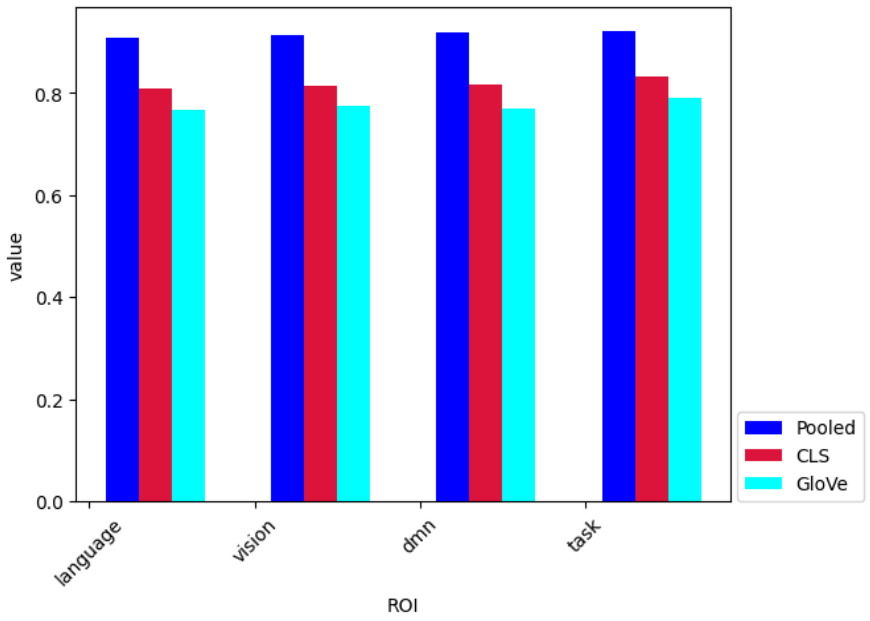
Therefore for brain encoder: GloVe > Pooled > CLS embeddings.

BRAIN DECODER

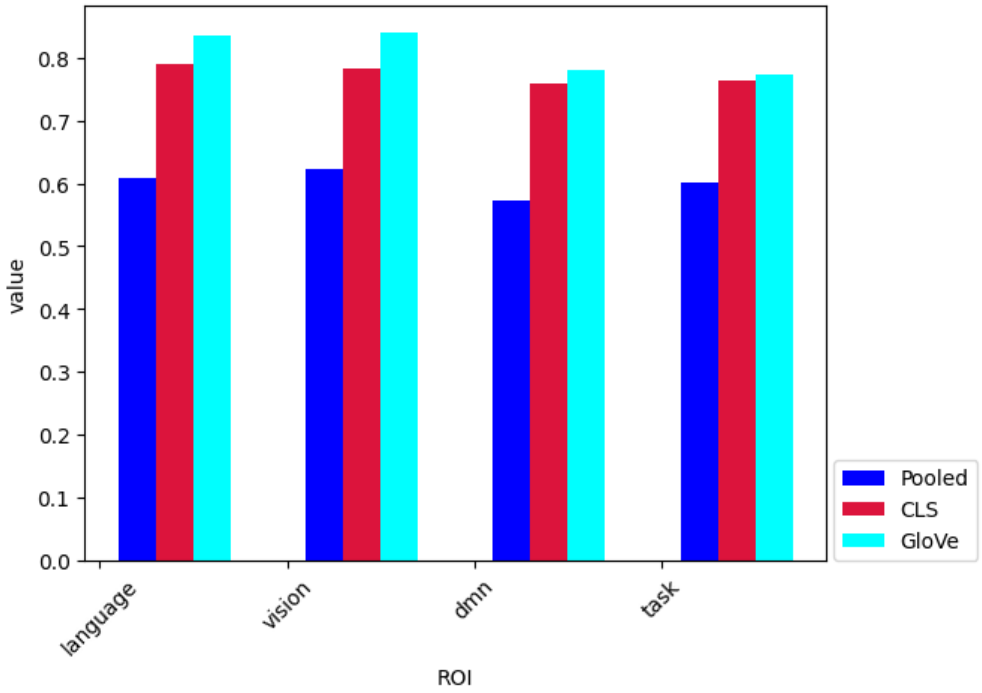
Subject 1 - 2v2 Accuracy



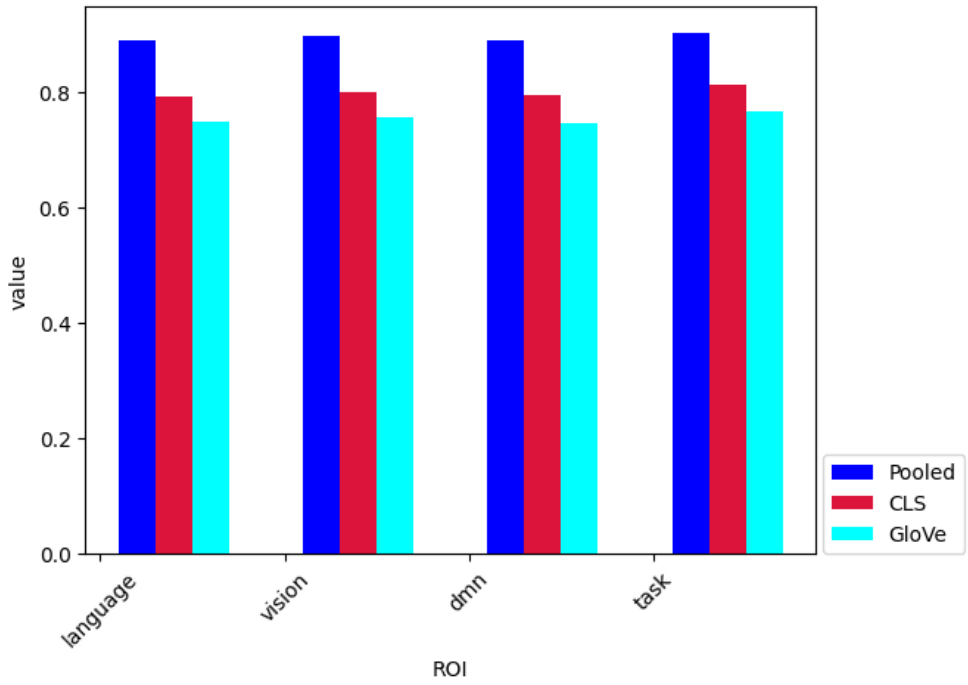
Subject 1 - Pearson correlation



Subject 2 - 2v2 Accuracy



Subject 2 - Pearson correlation



Based on the graphs presented above, it is evident that the expected sentence embeddings exhibit reasonable correlations when compared to the test data. The CLS embeddings demonstrate an exceptionally high score, indicating that the decoding model's performance was optimal when the sentence was embedded using CLS embeddings. Typically, CLS (and to some extent, Pooled) embeddings are anticipated to perform better when used in a decoding model as CLS embeddings capture the sentence's overall context. Conversely, when utilized as an encoder, CLS embeddings are expected to underperform. This is due to the alterations made while creating word embeddings, leading to information loss from the original sentence.

For decoding task also same for all ROIs:
BERT CLS > GloVe > BERT pooled for accuracy and GloVe < BERT CLS < BERT pooled for correlation.