

COMP 4475 Topics Artificial Intelligence

Project 2: Machine Learning in Julia (Word Embedding)

-Rakshit Saxena

0905094

In this project I have used Embeddings.jl to download pre trained Google Word2Vec word embeddings. I have used several packages like Distances.jl (for calculating cosine similarity), Unicode.jl (for removing punctuation marks from the corpus), Plots.jl (for plotting graphs). Closest function is used to display the words which are similar. I have used CorpusLoaders.jl for downloading the IMDB reviews which contain about 50000 reviews both negative and positive. Then to do the sentiment analysis I used Naïve Bayes Classifier with two classifiers being positive and negative. Then after I format the corpus into machine trainable form, using fit! I feed the model to the classifier. And finally using predict to predict the positive or negative sentiments for the sentences.