Lakshmi Kanikkannan

PH 1998L – Machine Learning in Practice

Homework 3

Sentiment Analyzer:

Code attached in: ml1998homework3.py

Excel sheet is attached

Chart, histogram

Description automatically generated

Fig 1 represents the distribution of the tweets of climate change and their respective polarity values.

The sentiments from these tweets are ever so slightly more positive than negative, but there’s a pretty even distribution on both sides.

ML Based Approach

Code attached in: ml1998homework3logreg.py

Text

Description automatically generated

Fig 2 shows the accuracy of these sentiment predictions on a test set fit in a logistic regression model. C refers to the relative strength of regularization, where smaller values of c correspond to stronger regularization, which involves adding a penalty to control the model complexity. With the highest level of regularization (c=0.75), the model is about 80.6% accurate, and even at the lowest level of regularization (c=10), the model is still 76.3% accurate, leading us to believe that the sentiment analyzer did relatively well at accurately predicting the sentiment of the tweet.

LDA

Code attached in: ml1998homework3lda.py

Excel output is also attached – the LDA successfully sorted tweets into various topics:

Text

Description automatically generated