LANL Earthquake Prediction Model

Team Name: Group 7

Course: CSCI 470: Machine Learning (OL Section), Canvas Group 7

Members: Elcin Eroglu, Ryan Sundberg, Linzhi Leiker, Cristian Madrazo

Apr 5, 2024

Proposal

Problem Statement

- The problem statement is predicting the time of the earthquake from real-time seismic data.
- Predicting earthquake timing accurately has many advantages. Urban planners can build resilient communities, insurance companies can improve risk assessment, government agencies can improve policy-making, scientific researchers can advance earthquake prediction techniques, emergency managers can improve preparedness, and infrastructure operators can protect assets.
- There are many studies using machine learning to predict earthquakes (Shreedharan et al., 2021). However, according to the USGS, there is no accurate way to predict the earthquake time currently (USGS, n.d.). So, the machine learning product that estimates the time of the earthquake can be an improvement to be prepared for the earthquakes and ensure the safety of the people.

Problem solution

- We plan to use a supervised machine learning model to use a regression model. A supervised regression model predicts continuous outcomes by learning the relationship between input features and target variables in labeled training data and identifying the underlying patterns.

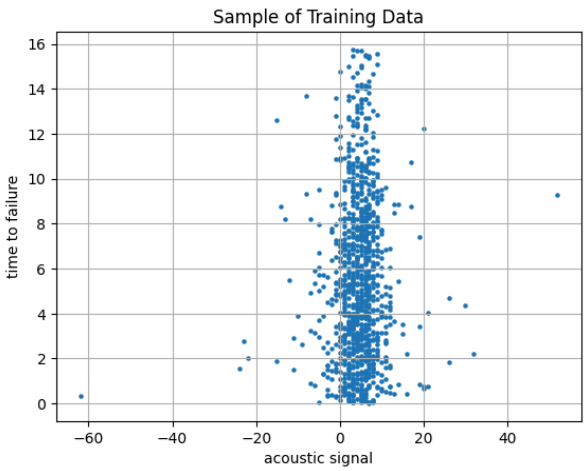
- The Root Mean Squared Error (RMSE) measure will be used to evaluate the resulting model. The RMSE metric is frequently utilized to evaluate the effectiveness of regression models.
-
- A 10% error on prediction time is acceptable. According to USGS, there is no approved method for predicting earthquakes in the US, therefore we believe that a prediction with standard 10% precision error is better than none.

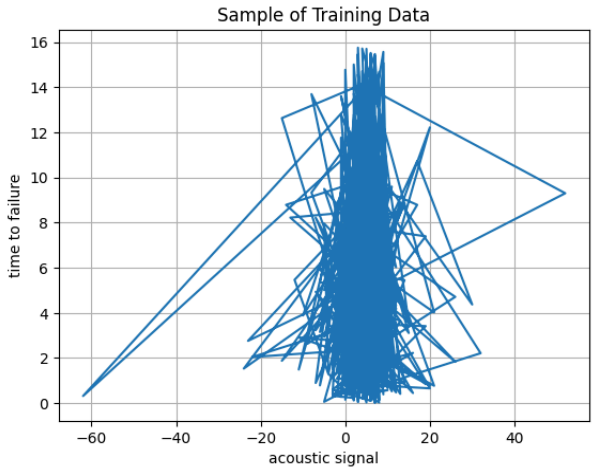
Data

- The training data we will be using for training the model is sourced from Los Alamos National Laboratory via Kaggle.
- The training data set is a 2 dimensional dataset with 6.291455e+08 data points.
- One dimension is the acoustic signal (an integer representing a signal from a seismometer).
- The other dimension is the time until the next earthquake occurs in seconds.
- The average data point has an acoustic signal of 4.52 and a time to failure value of 5.67s.
- The smallest time-to-failure is 9.550396e-05s while the biggest is 1.610740e+01s
- The smallest acoustic signal is -5.515000e+03 while the largest is 5.444000e+03

Additional Info

The first figure below shows a random sample (1000 points) of the training data. The second figure shows what a regression line may look like for that same data.





The table below shows a detailed description of the training data:

	acoustic_data	time_to_failure
count	6.29E+08	6.29E+08
mean	4.52E+00	5.68E+00
std	1.07E+01	3.67E+00
min	-5.52E+03	9.55E-05
25%	2.00E+00	2.63E+00
50%	5.00E+00	5.35E+00
75%	7.00E+00	8.17E+00
max	5.444000e+03	1.61E+01

Timeline

A timeline with **specific dates** for when we hope to **complete** project-relevant milestones:

- Construction of initial model (untrained) - April 5, 2024
- Training of initial model (including hyperparameter search) and assessment of model performance - April 12, 2024
- Training of revised model (or other model classes) and assessment of model performance - April 16, 2024
- Presenting final model and the results showing the estimated time to failure - April 24, 2024

References

Shreedharan, S., Bolton, D. C., Rivière, J., & Marone, C. (2021, July). *Machine learning predicts the timing and shear stress evolution of lab earthquakes using active seismic monitoring of Fault Zone Processes*. Journal of geophysical research. Solid earth. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9285915/

USGS. (n.d.). *Can you predict earthquakes?*. Can you predict earthquakes? | U.S. Geological Survey. https://www.usgs.gov/faqs/can-you-predict-earthquakes