

CS532 Project Proposal

Topic: Asteroids Classification

Dataset

The dataset to be used in this project is NASA Asteroids Classification dataset. (<https://www.kaggle.com/shrutimehta/nasa-asteroids-classification>). There are 4687 rows of data in total with 40 columns in each row. A total of 39 features represents the properties of the asteroid, such as the absolute magnitude, maximum estimated diameter, minimum estimated diameter, relative velocity, eccentricity and inclination. The label of each row is a binary class showing the asteroid is hazardous (True) or non-hazardous (False).

Two objectives will be realized. The major objective is to find the best classifier to classify the potential asteroids as hazardous or non-hazardous. The second objective is to determine the principle features that are responsible for claiming the asteroid to be hazardous.

Algorithm

Three algorithms will be applied on this dataset.

1. **Linear regression:** Least squares and ridge regression will be applied.
2. **Neural networks:** A simple two-layer artificial neural network will be built and trained.
3. **Support Vector Machine:** A kernel based SVM will be applied.

About 70% of the data will be grouped into training set and the rest 30% will be grouped into test set. A dimension reduction will be applied to select the principle features. When training with different algorithms, the k-fold cross-validation method will be used. To compare the performance of the three algorithms, a confusion matrix will be created to compare the numbers of the true labels and the predicted labels and the accuracy will be calculated. The algorithm with higher accuracy will be considered to have a better performance.

Timeline

Week	Tasks
10/19 – 10/25	Data preprocessing; feature selection.
10/26 – 11/01	Build, train and test linear classifier.
11/02 – 11/08	Optimize linear classifier; Build neural networks.
11/09 – 11/15	Train and test neural networks. First update 11/17.
11/16 – 11/22	Optimize neural networks. Build SVM.
11/23 – 11/29	Train, test and optimize SVM. Second update 12/1.
11/30 – 12/06	Validation and evaluation. Visualization and documentation.
12/07 – 12/13	Write project report 12/12.

Github Link:

<https://github.com/keshuw726/CS532-Course-Project-Asteroids-Classification>