STAT 432 Final Project

Detecting Volcanoes on Venus via Classification (Where are the Volcanoes?!!)

Team : Mamamia!
Di Ye (diye2), Hao Wang (haow2),Hannah Pu (chpu2)

November 17, 2018

Contents

Final Report

Introduction and literature review

Data Source information:

- The data was downloaded from Kaggle, which is originally from NASA's Magellan spacecraft database.
- Kaggle. https://www.kaggle.com/amantheroot/finding-volcanoes-on-venus/data

Data introduction and scientific goal:

9734 images were captured by the spacecraft and converted to pixel (110x110, from 0 to 255), where every image is one row of 12100 columns (all the 110 rows of 110 columns). Images can contain more than one volcanoes or maybe none. The 9000+ images are separated to four datasets (file names : *train_images*, *train_labels*, *test_images*, and *test_labels*):

(a) Image dataset (*train_images* and *test_images*) *Train_images* : 7000 images as train data with 12100 variables;
Test_images : 2734 images as test data with 12100 variables; All the variables correspond to the pixel image, 110 pixel * 110 pixel = 12100.

(b) Label dataset (*tain_labels* and *test_labels*) Both *train_labels* and *test_label* datasets include the following labels:
1. *Volcano?* : if in the image there are volcanoes (Main target), 1 (yes) or 0 (no)
(If Volcano? = 0, the following three categories would be "nan")
2. *Type* : 1= definitely a volcano,2 =probably, 3= possibly, 4= only a pit is visible
3. *Radius* : is the radius of the volcano in the center of the image, in pixels
4. *Number Volcanoes* : The number of volcanoes in the image
For this project, we will focus mainly on predicting whether each image has a volcanoe or not. In addition, if the classification prediction goes well, we will also construct model to predict the number of volcanoes in the images.