



“THE OUTLIERS” PRESENTS

Anomaly Detection Challenge -2 Satellite Image Data

NITHISH RAGHUNANDANAN

VISHAL BHALLA

Picture: Close-up of grandest canyon of all: Valles Marineris from ISRO's Mars Orbiter

Agenda

- ▶ Introduction
- ▶ Data Pre-processing
 - ▶ Rearrangement of the Data
 - ▶ Replacement of Missing Values
- ▶ System Pipeline
- ▶ Data Model
- ▶ Evaluation of the Models
- ▶ Results
- ▶ Conclusion

Introduction

- ▶ Aim: Classify the central pixel in 3x3 neighbourhoods of a multi-spectral satellite image.
- ▶ Problem Type: Multi-class Classification
- ▶ Samples
 - ▶ Training Set: 4435
 - ▶ Test Set: 2000
- ▶ Number of Attributes: 36 (4 spectral bands x 9 pixels in neighbourhood)
- ▶ Attributes: The attributes are numerical, in the range 0 to 255.
- ▶ Classification
 - ▶ There are 6 decision classes: 1,2,3,4,5 and 7.

Data Pre-processing

- ▶ Rearrangement of the Data
 - ▶ Given data set had the Spectral Data combined together.
 - ▶ Grouped all the data pertaining to a Spectrum together.
- ▶ Example

S1.1	S2.1	S3.1	S4.1	S1.2	S2.2	S3.2	S4.2
S1.1	S1.2	S2.1	S2.2	S3.1	S3.2	S4.1	S4.2

Data Pre-processing(2)

- ▶ Feature Analysis
 - ✓ Data is in the range of $[0,255]$
 - ✓ Normalisation not required
- ▶ Strategies for Replacement of Missing Values
 - ✓ Row Wise
 - ✓ Column Wise
 - ✓ Spectrum Wise

Data Pre-processing (3)

- ▶ Replacement Techniques
 - ✓ Zeros & Mid Values
 - ✓ Mean
 - ✓ Median
 - ✓ Mode
 - ✓ Minimum
 - ✓ Maximum
 - ✓ Interpolated Values

System Pipeline



- ▶ We iteratively used the above pipeline to arrive at the best model

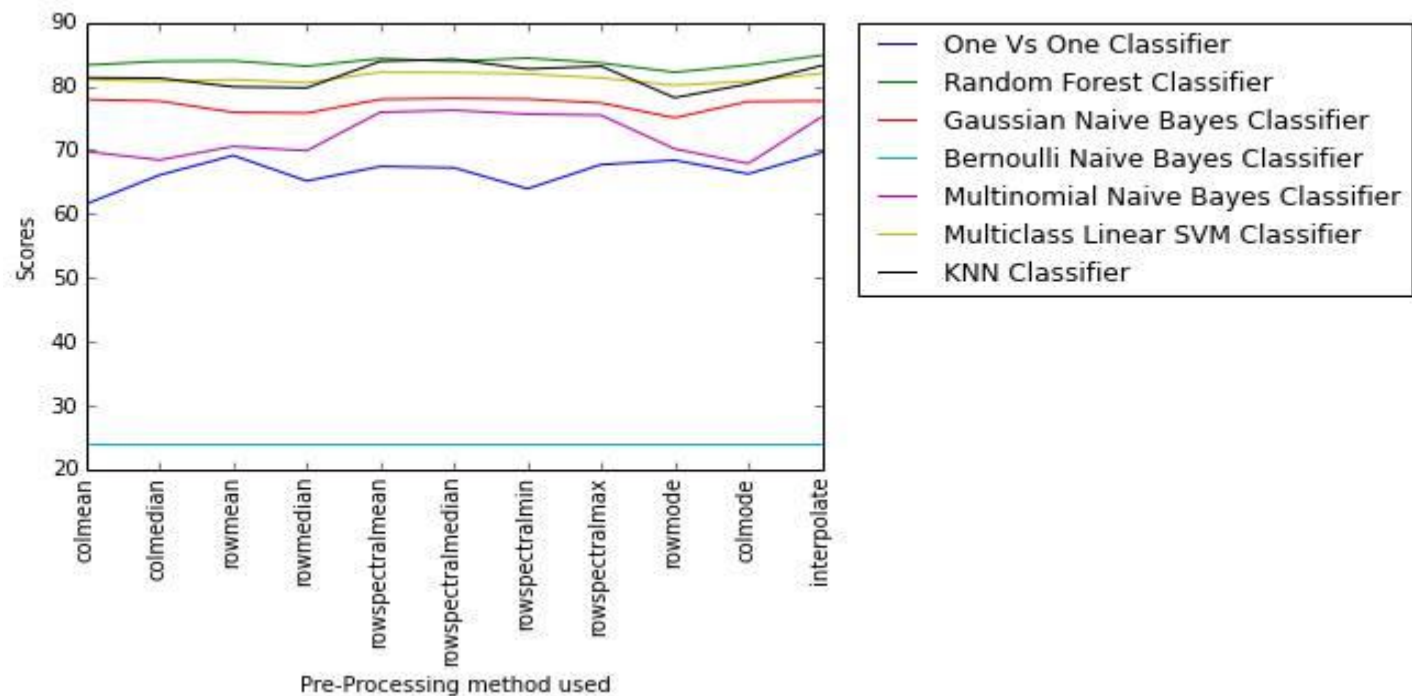
Model

- ▶ We tried different types of Multi Class Classification Models to fit our data
 - ▶ One Versus One and One Versus Rest Multiclass Classification
 - ▶ Random Forests
 - ▶ Gaussian , Bernoulli and Multinomial Naïve Bayes Classifier
 - ▶ Multi Class SVM with Linear Kernel
 - ▶ K Nearest Neighbours

Evaluation of the Models

Criteria

Classification Accuracy on Stratified K-Fold Cross Validation with a split of 4:1.



Results

Model	Replacement Technique	Classification Accuracy (Public Score)	Classification Accuracy (Private Score)
Random Forest (50 trees with Entropy)	Row Spectral Median	0.89800	0.92100
Random Forest (50 trees with Gini)	Row Spectral Median	0.89300	0.92400
KNN with 3 neighbours	Row Spectral Median	0.89200	0.90500
Random Forest(Gini)	Interpolation	0.89000	0.91800
KNN with 3 neighbours	Row Spectral Mean	0.89000	0.89700
KNN with 3 neighbours	Column Mean	0.87800	0.89900
Multi-class Linear SVM	Row Spectral Median	0.84000	0.85900

Conclusion

- ▶ The best results were observed for Random Forest (50 trees with Entropy as a criterion) on rearranged data and using row spectral median values for missing values.
- ▶ Replacement of missing values by Row Spectral Median values gave consistently better results across all of our models.



Questions?

Thank You !