Mini project : challenge Areal team sputnik

Members :

VERRIEZ CHASTANG Mathilde

DORDAIN Jean-Camille

HERBERA Mathis

Lydia ZENAIDI

Rayane DAHASSE

HUART Kevin

URL of the challenge :

https://codalab.lri.fr/competitions/399

Repo GitHub of project :

Context and description of the problem

* 13 categories

* 9100 satellite images

* size 32*32 pixels

* 700 images per class

* 2 parts : preprocess part and row data part

1) The preprocessing part

2) The prediction/classifier part

3) The vizualisation part

1) Preprocessing

* Load and transform raw data into understandable data

* 9100 images raw

* size 32*32 pixels

* algorithm of feature selection

* PCA

Feature selection (before)

0	0	1
0	1	1
0	0	0
0	1	1
1	1	0
0	1	1

Feature selection (after)

|	
0	1
1	1
0	0
1	1
1	0
1	1

2) Prediction

* Naive Bayes or Gaussian   Classifier

* SVM

* Decision Tree

* Random Forest

* Nearest Neighbors

Statistics and preliminary results

Method	NaiveBaiyes or Gaussian Classifier	SVM	Decision Tree	Random Forest	Nearest Neighbors
Training	0,81	0,81	0,57	0,64	0,92
CV	0,80	0,79	0,67	0,63	0,83
Validation	0,80	0,80	0,56	0,65	0,85

Preliminary results

3) The visualization part
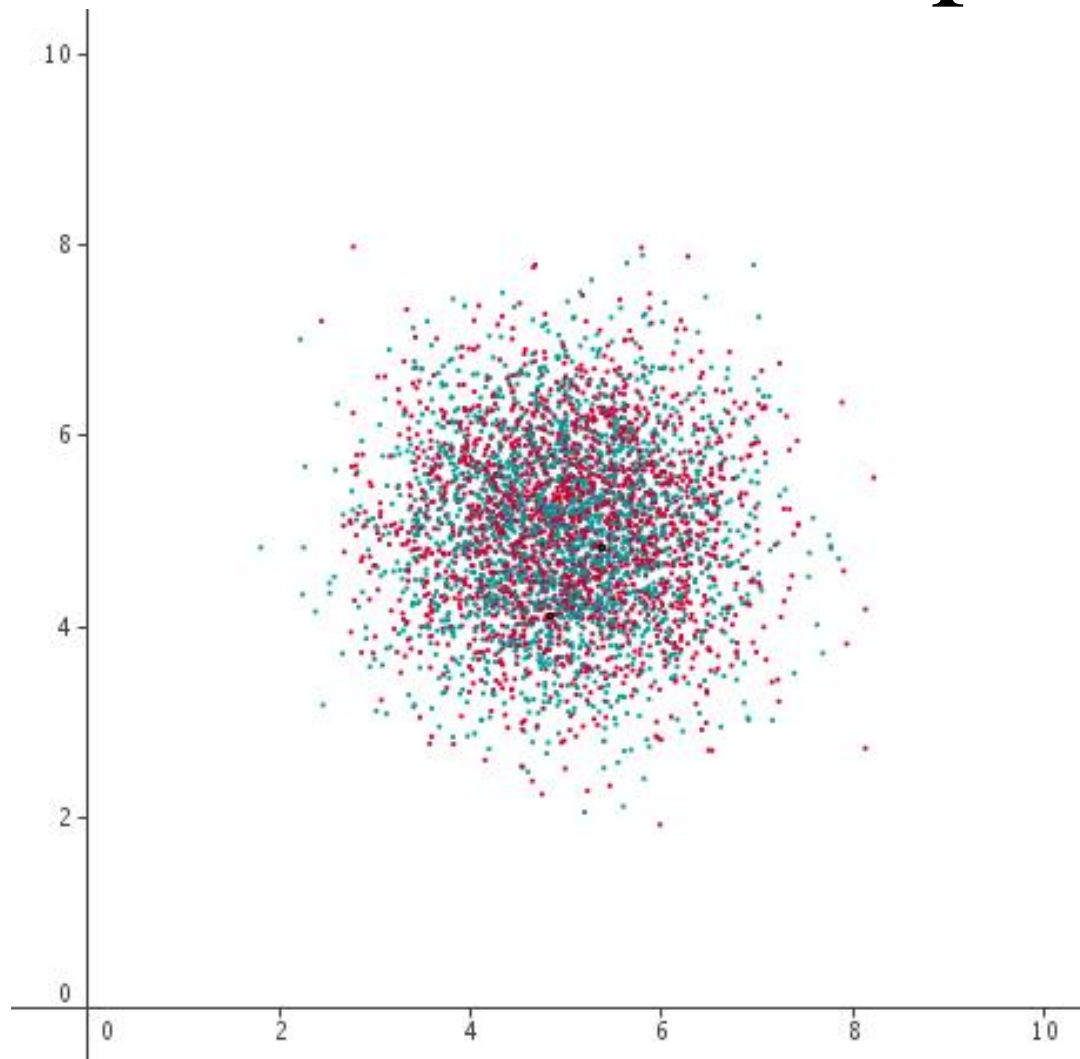
Point cloud

Conclusion

- Best score in Codalab, which is 0.899.

- Preprocessing part really important as we need to find the best balance between loss of data and the time it is going to take for the prediction part.