

# Anna\_Pavlenko\_Pixel\_Classification

May 6, 2020

## 1 Final Project: Earth Analytics Python Course: Spring 2020

## 2 Contents:

### 2.1 Earth Data Analytics Final Blog and Notebook

#### 2.1.1 Why Machine learning is important in Earth Science?

Recent intensive developments in Machine Learning (ML) have expanded application of artificial intelligence to different areas of our life: urban monitoring, fire detection or flood prediction (Fayyad et al., 1996.). The ML-based methods have been widely applied to the science and engineering problems for near two decades. This is while the application of these techniques in the geosciences and remote sensing area is fairly new and limited (David J.Lary, 2016).

Machine learning algorithms allowed the use of increasently available ‘big data’ like Remote Sensing: multispectral or radar satellite images, LiDAR high resolution data in automatization process of processing and preparing for future analyses.

A machine learning algorithm is a process that is used to fit a model to a dataset, through training or learning. The learned model is subsequently used against an independent dataset, in order to determine how well the learned model can generalise against the unseen data, a process called testing. In general, machine learning algorithms can be divided into two main groups (supervised- and unsupervised-learning; Fig. 1). Supervised-learning algorithms use predefined input-output pairs and learn how to derive outputs from inputs. The user specifies which variables (i.e., outputs) are considered dependent on others (i.e., inputs). The machine learning toolbox includes several linear and non-linear supervised learners, predicting either numeric outputs (regressors) or nominal outputs (classifiers) (Table 1). (SimonWillcock et al., 2018)

#### 2.1.2 Workflow:

- Machine learning processes automatically provide estimates of uncertainty.
- Uncertainty information enables decision-makers to assign their own thresholds.
- Machine learning algorithms can help scientists make use of ‘big data’.

Fig. 1. A schematic outlining how machine learning algorithms (yellow) can contribute to the data-driven modelling process (blue) (Fayyad et al., 1996).

## 2.2 Methods in this project

### 2.2.1 Functions

Table 1. A simplified summary of machine learning algorithms (categorised as supervised and unsupervised). (SimonWillcock et al., 2018)

### 2.2.2 Supervised Classification

**Creating ROI from.** First we need to define a region of interest (ROI). Instead of using an imported asset, we will use a single coordinate that we will manually define.

**Loading an ImageCollection and filtering to a single image** Now we will load Landsat imagery and filter to the area and dates of interest. We can use sort to filter the ImageCollection by % cloud cover, a property included with the Landsat Top of Atmosphere (TOA) collection. We then select the first (least cloudy) Image from the sorted ImageCollection // Load the Landsat 8 scaled radiance image collection. `var landsatCollection = ee.ImageCollection('LANDSAT/LC08/C01/T1').filterDate('2019-01-01', '2019-12-31');`

`// Make a cloud-free composite. var composite = ee.Algorithms.Landsat.simpleComposite({ collection: landsatCollection, asFloat: true });`

`// Visualize the Composite Map.addLayer(composite, {bands: ['B4', 'B3', 'B2'], max: 0.5, gamma: 2}, 'L8 Image', false);`

#### Collect Training Data from:

- coordinates
- manually collected points
- random points collection First we need to define a region of interest (ROI). Fig. 1. Random points collection

`// Merge points together var newfc = water.merge(urban).merge(forest); print(newfc, 'newfc')`

**Sample Imagery at Training Points to Create Training datasets** // Select the bands for training `var bands = ['B2', 'B3', 'B4', 'B5', 'B6', 'B7'];`

`// Sample the input imagery to get a FeatureCollection of training data. var training = composite.select(bands).sampleRegions({ collection: newfc, properties: ['landcover'], scale: 30 });`

**Train the classifier** // Make a Random Forest classifier and train it. `var classifier = ee.Classifier.randomForest().train({ features: training, classProperty: 'landcover', inputProperties: bands });`

**Classify the Image & Display the Results** Use the new classifier to classify the rest of the imagery.

Fig. 2 Supervised Classification from training data (MLCD was used)

**Assess the Accuracy** We can assess the accuracy of the trained classifier using a confusionMatrix.

```
// Get a confusion matrix representing resubstitution accuracy. print('RF error matrix:', classifier.confusionMatrix()); print('RF accuracy:', classifier.confusionMatrix().accuracy());
```

### 2.2.3 Unsupervised Classification

Need a accuracy control/ validation of result.

### 2.2.4 Regression spectral un-mixing

```
// Instantiate the clusterer and train it. var clusterer = ee.Clusterer.wekaKMeans(15).train(training);  
// Cluster the input using the trained clusterer. var result = selection.cluster(clusterer);  
// Display the clusters with random colors. Map.addLayer(result.randomVisualizer(), {}, 'clusters');
```

### 2.2.5 Un-mixed result:

Need a accuracy control/ validation of result.

## 3 Result

For the ML need to create more training points from high resolution imagery (NAIP imagery) and combine NDVI.

## 4 Summary / Conclusions:

### 4.1 Development in the future

- 1) Consider 1 Year Time Frame for Landsat / Tropi Data in EE.
- 2) Set a Training Data Set from NAIP imagery (high resolution)
- 3) Work on Deep Machine Learning - TensorFlow model

## 4.2 Long-term goal:

Build Machine Deep Learning Project:

- Creating a TensorFlow Deep Learning VM Instance
- Accuracy validation the TensorFlow compared to Supervised/ Unsupervised Classification

## 5 References:

- 1) Fayyad et al., 1996. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth. “From data mining to knowledge discovery in databases”. AI Mag., 17 (1996), p. 37, 10.1609/AIMAG.V17I3.1230
- 2) David J.Lary, “Machine learning in geosciences and remote sensing”. Geoscience Frontiers. Volume 7, Issue 1, January 2016, Pages 3-10
- 3) SimonWillcock et al. “Machine learning for ecosystem services”. Ecosystem Services Volume 33, Part B, October 2018, Pages 165-174. <https://doi.org/10.1016/j.ecoser.2018.04.004>
- 4) LEO BREIMAN Statistics Department, University of California, Berkeley, CA 94720. “Random Forests”. Machine Learning, 45, 5–32, 2001 Kluwer Academic Publishers. Manufactured in The Netherlands

[ ]: