

ENVI Machine Learning Tutorial: Unsupervised Classification

In this tutorial you will use ENVI Machine Learning to train an unsupervised model, perform classification, and display the results. Unsupervised classifiers do not require labeled data to identify features in imagery. These types of classifiers identify features based on the number of classes requested.

System Requirements

Refer to the *Systems Requirements* topic in the ENVI Deep Learning Help.

The steps for this tutorial were successfully tested with ENVI Deep Learning 2.0, the module version that provides ENVI Machine Learning. An Intel® Xeon® W-10855M CPU @ 2.80GHz 2.81GHz was used to run this tutorial (not required). Intel CPUs are recommended for performance gains due to the use of Intel libraries (not required).

Files Used in This Tutorial

Sample data files are available on our [ENVI Tutorials](#) web page. Click the "Deep Learning" link in the ENVI Tutorial Data section to download a .zip file containing the data. Extract the contents to a local directory. Files are located in the `machine_learning\unsupervised` folder.

The image used for this tutorial was collected by the James Webb Space Telescope (JWST) (<https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html>) NASA JPL. Images are in the public domain. License terms are governed by CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0>).

The image is of a spiral galaxy NGC 628, also known as Messier 74, located 32 million light-years from Earth. This is a three-band image constructed from three single band rasters, and saved in the ENVI file format.

File	Description
JWST_ngc628.dat	JWST image (1,977 x 1,149 pixels) used for training and classification.

Background

Unsupervised learning is helpful for data science teams who are unfamiliar with a particular scene and its features. This has the advantage of categorizing unknown similarities and differences in the data by grouping alike features. Having a good sense of what might be in the

data can help fine tune results. The number of classes requested before training determines the number of output classes/features.

Advantages of Unsupervised Learning

- Labeled data is not required. This saves time by only specifying the raw imagery containing possible features of interest.
- Simpler and faster to acquire unlabeled data.
- Potential to identify features or patterns not easily noticeable by the human eye.
- Reduces the chance of human error bias; this can happen while manually labeling data for a supervised approach.

Disadvantages of Unsupervised Learning

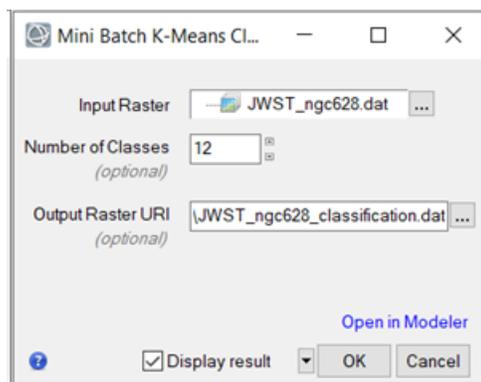
- The user is responsible for adding meaning to features or patterns identified by the classifier, as opposed to supervised learning, where all labels are known ahead of time.
- Requesting many features can saturate classification results, while requesting not enough features can mask information that might have been visible if more features were requested.

Train and Classify an Unsupervised Classifier

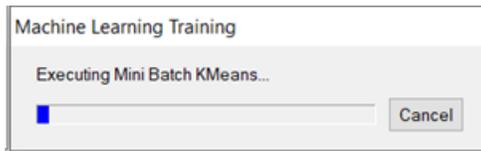
ENVI Machine Learning provides several different ways to train and classify data. For this tutorial we will use the Mini Batch K-Means Classification task, which will perform training and classification with a single raster. Using a single raster will produce a model specific to the image used for training. As a result, this model will likely not produce usable classification results when applied to other imagery. To create a generalized model, multiple rasters are required, this is possible using the ENVITask API or the ENVI Modeler. See these topics in ENVI Help for details.

The MiniBatchKMeans is a variant of the KMeans algorithm which uses mini-batches to reduce the computation time, while still attempting to optimize the same objective function. Mini-batches are subsets of the input data, randomly sampled in each training iteration. These mini-batches drastically reduce the amount of computation required to converge to a local solution. In contrast to other algorithms that reduce the convergence time of k-means, mini-batch k-means produces results that are generally only slightly worse than the standard algorithm. (https://scikit-learn.org/stable/modules/clustering.html?force_isolation=true#mini-batch-kmeans)

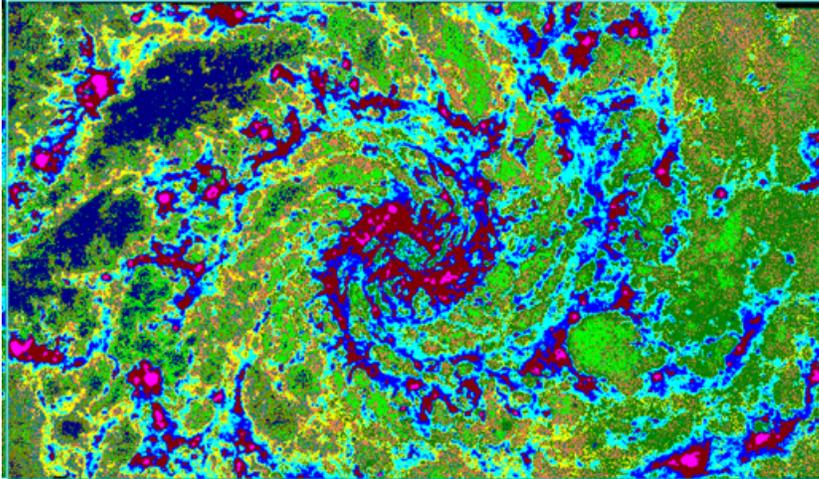
1. Start ENVI
2. In the ENVI Toolbox, expand the **Machine Learning** folder and select **Classification > Unsupervised > Mini Batch K-Means Classification**. The Mini Batch K-Means Classification dialog appears.
3. Click the **Browse** button  next to **Input Raster**. The Data Selection dialog appears.
4. Click the **Open File** button  at the bottom of the Data Selection dialog.
5. Go to the directory where you saved the tutorial data, navigate to `machine_learning\unsupervised`, select `JWST_ngc628.dat`, and click **Open**.
6. With the `JWST_ngc628.dat` file selected, click **OK**.
7. Optionally, change the default **Number of Classes** from 3 to 12 by toggling the up/down arrows, or enter a new value in the field.
8. Optionally, specify the output location for the **Output Raster**.
 - a. Click the **Browse** button  next to **Output Raster**. The Select Output Raster dialog appears.
 - b. Specify an output directory location by setting the folder path at the top of the dialog, then press **Enter**.
 - c. Enter the **File Name** `JWST_ngc629_classification.dat` and click **Open**.
9. Leave the **Display result** check box selected, and click **OK**.



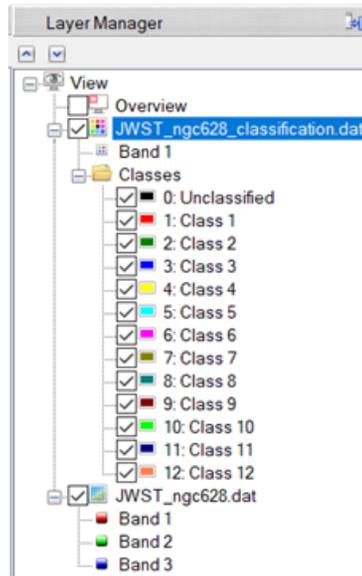
10. The training process begins and the Machine Learning Training progress dialog appears. Training takes seconds to complete.



11. When training completes, the `JWST_ngc628_classification.dat` image is displayed in ENVI. Due to the nature of Machine Learning randomness, results may vary. There is no guarantee your trained model will produce an identical result to the one below.



12. Examine identified features by overlaying the `JWST_ngc628_classification.dat` image over the original raster `JWST_ngc628.dat`. Use ENVI's **Transparency** tool to identify features by color classified by the unsupervised model.
 - a. Load the Input Raster `JWST_ngc628.dat` into ENVI's display, press **F4**, and the Data Manager dialog appears. Right-click **JWST_ngc628.dat** > **Load Default**. The raster `JWST_ngc628.dat` appears in the Layer Manager and is displayed.
 - b. In the Layer Manager, click and hold `JWST_ngc628_classification.dat` and drag it above `JWST_ngc628.dat` in the list. The classification raster is now overlaying the input raster.
 Done correctly, the Layer Manager should look like the following image.



Use ENVI's **Transparency** tool to toggle back and forth. Notice that variations of gas clouds and features identified by the unsupervised classifier.

13. When you are finished, exit ENVI.

Final Comments

In this tutorial, you learned how to use ENVI Machine Learning to extract twelve features from a galaxy 32 million light years away from Earth. You learned a traditional machine learning approach to unsupervised classification using Mini Batch K-Means Classification.

It is recommended that you experiment with the number of classes as the results can drastically change. As mentioned previously, the model that is created in this tutorial is not generalized. Meaning, it is not likely to produce a useful classification result when used on different rasters. If you are interested in producing a more general purpose model, refer to other classification topics ENVI Machine Learning Help.

In conclusion, traditional machine learning approaches provide a simple yet powerful way to visualize data. Machine learning with regards to imagery only requires hundreds to thousands of pixels in order to identify patterns and features in your data.

For more information about the capabilities presented here, or other machine learning approaches such as supervised classification, and anomaly detection, refer to the *Overview of ENVI Machine Learning* topic in ENVI Machine Learning Help.