Machine Learning Module Descriptions

Updated: June 22, 2015

The machine learning libraries provided in Azure Machine Learning Studio are encapsulated in modules, each performing a task common in machine learning scenarios. By connecting and then configuring modules, you can create a workflow that reads data from external sources, prepares it for analysis, applies machine learning algorithms, and generates results.

For more information about how to drag and drop modules to create a machine learning experiment, see these tutorials:

- Develop a predictive solution by using Azure Machine Learning (http://azure.microsoft.com/documentation/articles/machine-learning-walkthrough-develop-predictive-solution/)
- Create a simple experiment in Azure Machine Learning Studio (http://azure.microsoft.com/documentation/articles/machine-learning-create-experiment/)

Module Categories

The machine learning tools or modules in Azure Machine Learning Studio are grouped by the following categories.

Category
Data Format conversions (https://msdn.microsoft.com/en-us/library/azure/dn906017.aspx) Use these modules to convert data to one of the formats used by other machine learning tools or formats.
Data Input and Output (https://msdn.microsoft.com/en-us/library/azure/dn906024.aspx) Use these modules to read data from cloud data sources, including Hadoop clusters, Azure table storage, and Web URLs.
Data Transformation (https://msdn.microsoft.com/en-us/library/azure/dn905834.aspx) Use these modules to prepare data for analysis. You can change data types, flag columns as features or labels, generate features, and scale or normalize data, and much more. • Filter (https://msdn.microsoft.com/en-us/library/azure/dn906005.aspx) Transform numeric data derived from digital signal processing.

- Learning With Counts (https://msdn.microsoft.com/en-us/library/azure/dn913056.aspx)

  Use joint probability distributions to build features that compactly describe large datasets.
- Manipulation (https://msdn.microsoft.com/en-us/library/azure/dn905863.aspx)

  This group provides a variety of tools for data science: remove or replace missing values, choose a subset of columns, add column or concatenate two datasets, and so forth.
- Sample and Split (https://msdn.microsoft.com/en-us/library/azure/dn905973.aspx)

  Divide a dataset by criteria or by size, to create training and test sets, or to isolate certain rows.
- Scale and Reduce (https://msdn.microsoft.com/en-us/library/azure/dn905911.aspx)

  Transform numerical data.

Feature Selection (https://msdn.microsoft.com/en-us/library/azure/dn905912.aspx)

Use these modules to identify the best features in your data, using widely researched statistical methods.

Machine Learning (https://msdn.microsoft.com/en-us/library/azure/dn905870.aspx)

This group contains most of the machine learning algorithms supported by Azure Machine Learning.

It also contains modules intended to support the algorithms by training models, generating scores, and evaluating model performance.

- Evaluate (https://msdn.microsoft.com/en-us/library/azure/dn906026.aspx)

  After you have trained a model, use these tools to measure the model's accuracy.
- Initialize (https://msdn.microsoft.com/en-us/library/azure/dn905812.aspx)

  These modules provide the machine learning algorithms, which you can customize by setting parameters. The algorithms in this section are grouped by type:

  - Anomaly detection algorithms (https://msdn.microsoft.com/en-us/library/azure/dn913096.aspx)
  - Classification algorithms (https://msdn.microsoft.com/en-us/library/azure/dn905808.aspx)
  - Clustering algorithms (https://msdn.microsoft.com/en-us/library/azure/dn905908.aspx)
  - Regression algorithms (https://msdn.microsoft.com/en-us/library/azure/dn905922.aspx)
- Score (https://msdn.microsoft.com/en-us/library/azure/dn906012.aspx)

  Use these modules to pass new data through the algorithm and generate a set of results for evaluation.

You can also use the results of scoring as part of a predictive service. • Train (https://msdn.microsoft.com/en-us/library/azure/dn905846.aspx) These modules train an initialized machine learning model on data you provide.
OpenCV Library Modules (https://msdn.microsoft.com/en-us/library/azure/dn905946.aspx) These modules give you easy access to a popular open source library for image processing and image classification.
R Language Modules (https://msdn.microsoft.com/en-us/library/azure/dn905920.aspx) Use these modules to add custom R code to your experiment, or implement a machine learning model based on an R package.
Python Language Modules (https://msdn.microsoft.com/en-us/library/azure/dn927167.aspx) Use these modules to add custom Python code to your experiment.
Statistical Functions (https://msdn.microsoft.com/en-us/library/azure/dn905867.aspx) Use these modules to calculate probability distributions, create custom calculations, and perform a wide variety of other tasks related to numerical variables.
Text Analytics (https://msdn.microsoft.com/en-us/library/azure/dn905886.aspx) Use these modules to perform feature hashing and named entity recognition.

See Also

A-Z List of Machine Learning Studio Modules (https://msdn.microsoft.com/en-us/library/azure/dn906033.aspx)