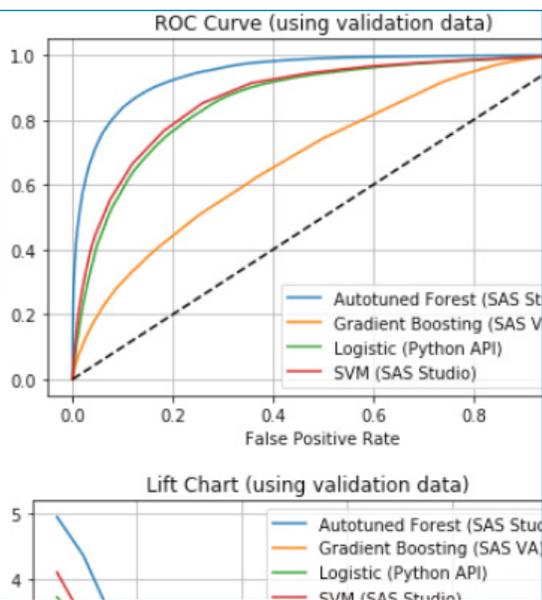


SAS[®] Machine Learning on SAS[®] Analytics Cloud

On-demand programming access to machine learning algorithms in the cloud



SAS Analytics Cloud is a powerful, new way for SAS to provision and deliver world-class software as a service (SaaS) to our customers. Running on the SAS Cloud and using the latest container technology, SAS Analytics Cloud simplifies access to our applications by eliminating the need to install, update or maintain software or infrastructure. Customers can simply sign up, sign in and get to work.

SAS Machine Learning on SAS Analytics Cloud is designed for SAS and open source data scientists to gain on-demand programmatic access to SAS Viya. It includes all the algorithms that sit behind SAS Visual Data Mining and Machine Learning, SAS Visual Statistics and SAS Visual Analytics. This means:

- No software to install or maintain.
- No hardware or infrastructure to buy or maintain.
- Users can manage and collaborate with others.
- Users can upload their own data.
- Affordable user-based licensing.

What does SAS[®] Machine Learning on SAS[®] Analytics Cloud do?

SAS Machine Learning combines data preparation, feature engineering, modern statistical and machine learning techniques in a single, scalable in-memory processing environment to develop, test and deploy models. It supports the entire machine learning process, from data transformation and preparation to scoring and deploying, in a single environment. It also supports the popular open source language Python and the SAS language.

Why is SAS[®] Machine Learning on SAS[®] Analytics Cloud important?

Simplified user-based licensing and a credit card payment means no software to install. It's the simplest, faster way to get started using SAS. Data scientists will benefit by having a collaborative coding environment that serves both SAS and Python programmers so they can instantly run some of the most advanced machine learning algorithms in the world.

For whom is SAS[®] Machine Learning on SAS[®] Analytics Cloud designed?

It is designed to support data scientists who prefer accessing SAS[®] Viya[®] through a programming interface using either the SAS or Python programming languages. These capabilities empower data scientists to analyze large and complex data, and build predictive models using a modern, powerful and customizable programming language.

Benefits

- **Quickly identify patterns, trends and relationships in structured and unstructured data.** With no software to install and simplified, user-based licensing, SAS Machine Learning on SAS Analytics Cloud provides fast, easy access to a broad set of modern statistical, machine learning, deep learning and text analytics algorithms. It also provides advanced data transformation and manipulation.
- **Provide easy, cloud-based access to data science applications.** Give data scientists instant access to powerful data science applications via SAS Analytics Cloud. We take care of the infrastructure, setup and maintenance so everyone else can focus on using the applications.
- **Reduce learning curves and expedite model development.** Boost programming efficiency with autocomplete capabilities, syntax help, color coding, keyword shortcuts and more. Automatic code generation and reusable code snippets help your programmers code machine learning algorithms faster.
- **Create better models.** Using advanced optimization techniques, an integrated autotuning process adjusts large combinations of hyperparameter settings and returns the optimal set automatically for the data scientist. Best-in-class hyperparameter autotuning enables faster, better modeling results.
- **Collaborate across languages.** SAS Machine Learning lets data scientists code in their preferred programming language. In addition to using the SAS language, SAS algorithms can be accessed from Jupyter Notebook using open source Python. Data scientists with varied programmatic skill sets can all contribute, taking advantage of existing personnel knowledge with a native interface to Python and the SAS programming language.
- **Easily navigate the entire analytics life cycle.** A single programming interface facilitates fast, efficient movement from data preparation, to exploration, to model development.

Overview

SAS Machine Learning on SAS Analytics Cloud combines data preparation, feature engineering, modern statistical and machine learning techniques in a single, scalable in-memory processing environment so data scientists can quickly develop, test and deploy models.

Machine learning algorithms

SAS Machine Learning on SAS Analytics Cloud delivers a broad set of modern statistical, machine learning, deep learning and text analytics algorithms within a single environment. This includes neural networks,

clustering, different flavors of regression, forest, gradient boosting models, support vector machines, language processing, topic detection and more.

These powerful methods drive the identification of new patterns, trends and relationships between data attributes in structured and unstructured data. The solution also provides matrix factorization for building customized recommendation systems.

Deep learning algorithms include deep neural networks, convolution neural networks for image classification and

Key Features

Interactive programming in a web-based development environment

- Code generator expedites coding for machine learning algorithms.
- Reusable code snippets.

Model development with modern machine learning algorithms

- Decision forests:
 - Distribution of independent training runs.
 - Supports intelligent hyperparameter autotuning of model parameters.
 - Generation of SAS ASTOREs for production scoring.
- Gradient boosting:
 - Automated iterative search for optimal partition of the data in relation to selected label variable.
 - Automated resampling of input data with adjusted weights based on residuals.
 - Automated generation of weighted average for final supervised model.
 - Supports binary, nominal and interval labels.
 - Ability to customize tree training with variety of options for numbers of trees to grow, splitting criteria to apply, depth of subtrees and compute resources.
 - Automated stopping criteria based on validation data scoring to avoid overfitting.
 - Generation of SAS ASTOREs for production scoring.
- Neural networks:
 - Automated intelligent tuning of parameter set to identify optimal model.
 - Supports modeling of count data.
 - Intelligent defaults for most neural network parameters.
 - Ability to customize neural networks architecture and weights.
 - Techniques include deep forward neural network (DNN), convolutional neural networks (CNNs), recurrent neural networks (RNNs) and autoencoders.
 - Ability to use an arbitrary number of hidden layers to support deep learning.

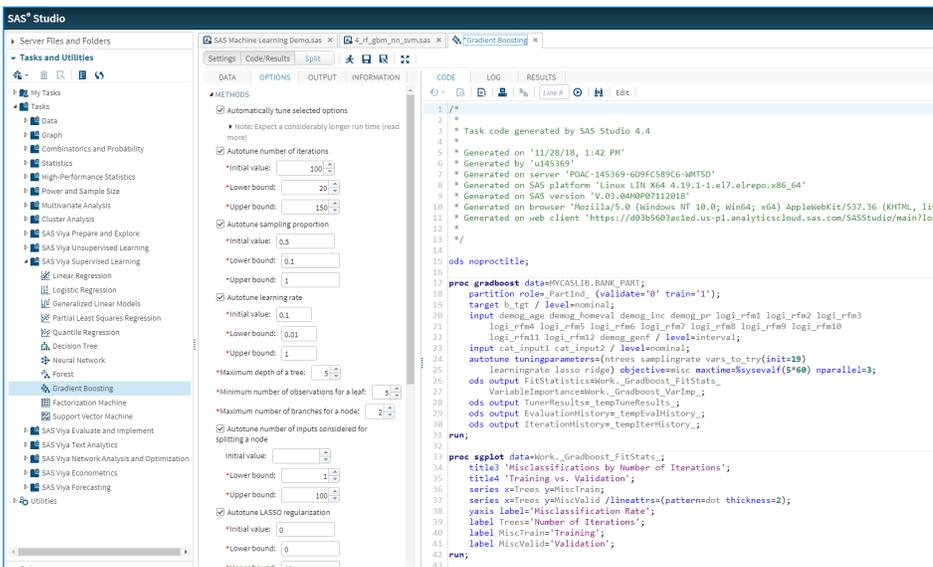


Figure 1: SAS Machine Learning on SAS Analytics Cloud delivers a broad set of modern statistical, machine learning, deep learning and text analytics algorithms.

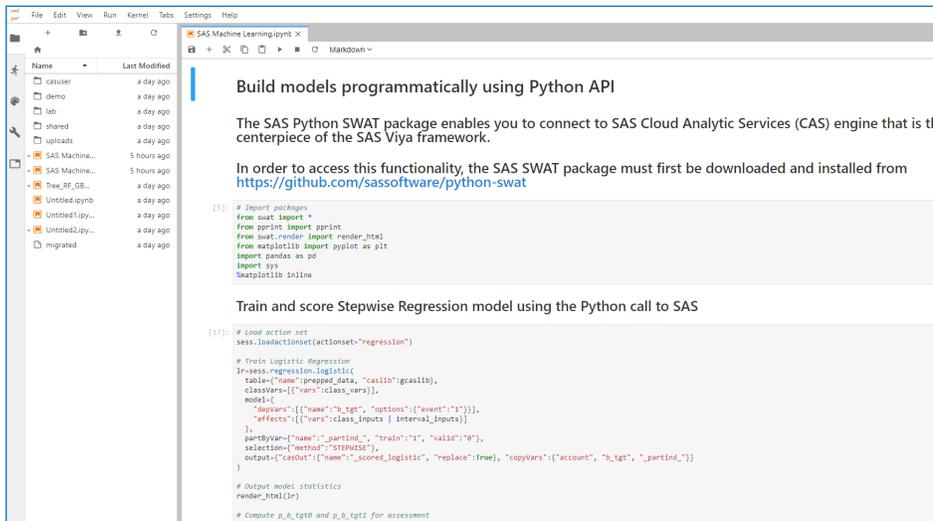


Figure 2: Data scientists can access SAS algorithms from Jupyter using Python.

recurrent neural networks for improved text analysis. Users can customize these networks and support different types of layers such as convolution and pooling.

Autotuning and optimization solvers

Use SAS autotuning to automate the process of finding the best set of machine learning hyperparameters or properties based on your modeling objective. Take advantage of built-in optimization solvers to build the most optimal models, in a short amount of time. Spend more time on generating meaningful insights, and let SAS crunch the numbers for you.

Access SAS® from open source languages

In addition to the SAS language, data scientists can access SAS algorithms from Jupyter Notebook using open source Python. Generate models, assess output and generate insights within this unified experience. Access the same machine learning algorithms and data that is available via SAS programming in SAS Studio.

Highly scalable, in-memory analytical processing

This solution provides a multithreaded, multiuser environment for concurrent access to data in memory. Data and analytical workload operations are multithreaded and automatically operate in parallel to produce faster results.

All data, tables and objects are held in memory as long as required, allowing for efficient processing. With built-in fault tolerance and memory management, advanced workflows can be applied to data, ensuring that processes always finish.

You get dramatically reduced runtimes for large data and analytical processing, reduced network traffic and can take full advantage of modern, multicore architectures to find solutions much faster.

Key Features (continued)

- Automatic standardization of input and target variables.
- Automatic out-of-bag validation for early stopping to avoid overfitting.
- Supports intelligent hyperparameter autotuning of model parameters.
- Generation of SAS ASTOREs for production scoring.
- Support vector machines:
 - Models binary target labels.
 - Supports linear and polynomial kernels for model training.
 - Ability to include continuous and categorical in/out features.
 - Automated scaling of input features.
 - Ability to apply the interior-point method and the active-set method.
 - Supports data partition for model validation.
 - Supports cross-validation for penalty selection.
 - Generation of SAS ASTOREs for production scoring.
- Factorization machines:
 - Supports the development of recommender systems based on sparse matrices of user IDs and item ratings.
 - Ability to apply full pairwise-interaction tensor factorization.
 - Includes additional categorical and numerical input features for more accurate models.
 - Supercharge models with timestamps, demographic data and context information.
 - Supports warm restart (update models with new transactions without full retraining).
 - Generation of SAS ASTOREs for production scoring.
- Bayesian networks:
 - Learns different Bayesian network structures, including naive, tree-augmented naive (TAN), Bayesian network-augmented naive (BAN), parent-child Bayesian networks and Markov blanket.
 - Performs efficient variable selection through independence tests.
 - Selects the best model automatically from specified parameters.
 - Generation of SAS ASTOREs for production scoring.
- Dirichlet Gaussian mixture models (GMMs):
 - Can execute clustering in parallel and is highly multithreaded.
 - Performs soft clustering, which provides not only the predicted cluster score but also the probability distribution over the clusters for each observation.
 - Learns the best number of clusters during the clustering process.
 - Uses a parallel variational Bayes (VB) method as the model inference method, which approximates the (intractable) posterior distribution and then iteratively updates the model parameters until it reaches convergence.
- Semisupervised learning algorithm:
 - Highly distributed and multithreaded.
 - Returns the predicted labels for both the unlabeled data table and the labeled data table.
- T-distributed stochastic neighbor embedding (t-SNE):
 - Highly distributed and multithreaded.
 - Returns low-dimensional embeddings that are based on a parallel implementation of the t-SNE algorithm.

Statistical algorithms

- Clustering (k-means, k-modes or k-prototypes clustering).
- Decision trees.
- Logistic regression.
- Linear regression.
- Generalized linear models.
- Generalized additive models.
- Nonparametric logistic regression.
- Market basket analysis.
- k-NN (k-nearest neighbor).

Integrated text analytics

Integrated text analytics are included for users who want to incorporate features derived from free-form text into a predictive model for text parsing and topic discovery, automatic Boolean-rule generation for categorical target variables and scoring data for text topics.

Combining structured data with text data uncovers previously undetected relationships and adds even more predictive power to analytical models.

Model assessment and scoring

Now it's easy to test different modeling approaches in a single run, and compare results of multiple machine learning algorithms with standardized tests to automatically identify champion models.

TO LEARN MORE »

To learn more about SAS Machine Learning on SAS Analytics Cloud and see other related material, please visit sas.com/ml-cloud.

Key Features (continued)

Analytical data preparation

- Distributed data management routines provided via code:
 - T-distributed stochastic neighbor embedding (t-SNE).
 - Feature binning.
 - High-performance imputation of missing values in features with user-specified values, mean, pseudo median and random value of non-missing values.
 - Feature dimension reduction.
 - Large-scale principal components analysis (PCA), including moving windows and robust PCA.
 - Unsupervised learning with cluster analysis and mixed variable clustering.
- Large-scale data exploration and summarization.
- Large-scale data profiling of input data sources.
- Sampling: Supports random and stratified sampling, oversampling for rare events and indicator variables for sampled records.

Integrated text analytics

- Supports 32 native languages out of the box: English, Arabic, Chinese, Croatian, Czech, Danish, Dutch, Farsi, Finnish, French, German, Greek, Hebrew, Hindi, Hungarian, Indonesian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Slovenian, Spanish, Swedish, Tagalog, Turkish, Thai and Vietnamese.
- Automated parsing, tokenization, part-of-speech tagging and lemmatization.
- Predefined concepts extract common entities such as names, dates, currency values, measurements, people, places and more.
- Automated feature extraction with machine-generated topics (singular value decomposition and latent Dirichlet allocation).
- Supports machine learning and rules-based approaches within a single project.
- Automatic rule generation with the BoolRule.
- Classify documents more accurately with deep learning (recurrent neural networks).

Model assessment

- Automatically calculates supervised learning model performance statistics.
- Produces output statistics for interval and categorical targets.
- Creates lift table for interval and categorical target.
- Creates ROC table for categorical target.

Model scoring

- Automatically generates SAS DATA step code for model scoring.
- Applies scoring logic to training, holdout data and new data.

SAS® procedures (PROCs) and SAS® Cloud Analytic Services (CAS) actions

- A programming interface (SAS Studio) allows IT or developers to access a CAS server, load and save data directly from a CAS server, and support local and remote processing on a CAS server.
- Python programmers or IT staff can access data and perform basic data manipulation against a CAS server or execute CAS actions using PROC CAS.

To contact your local SAS office, please visit: sas.com/offices

