



Nyheder i SAS' analytiske produkter

Analytics netværksmøde i SAS 26. maj 2020

SAS Viya 3.5 – Analytiske nyheder

Produktivitet/ Automatisering

- Automatiseret machine learning
 - Hyperparameter tuning
 - Forslag til dannelse af features
 - Automatiseret dannelse af model pipeline

Open Source integration

- Open Source Model Manager

Fortolkbarhed

- Udover LIME, ICE og PD plots er der kommet Kernel SHAP
- Fortolkning af Cumulative Lift, ROC mv. nedfældet i tekst

Visual Statistics 8.5

SAS Visual Statistics 8.5 (November 2019 release of SAS Viya 3.5) includes these new features and enhancements to the user interface:

- A confusion matrix that displays the classification results for categorical response models is available.
- Several new assessment statistics are available for all models.
 - You can now derive predicted items from generalized additive models and nonparametric logistic regression models.
 - You can now export and save generalized additive models and nonparametric logistic regression models.
 - In the Cluster object, a details table has been added for the Parallel Coordinates plot.
 - The **Bin method** property that specifies the way that measure predictors are binned has been added to the Decision tree object.
 - There is a new option for autotuning decision tree models: **Leaf size**.
 - In the settings, you can now specify to automatically convert measure variables with two levels to a category when a data set is first opened in SAS Visual Analytics.
 - In the settings, you can now set a default assessment statistic to use in all models.

SAS Visual Data Mining and Machine Learning 8.5

- The new KPCA procedure performs kernel principal component analysis (KPCA).

KPCA is a nonlinear extension to the widely used principal component analysis (PCA). It uses the "kernel trick" to implicitly map the original data to some high-dimensional RKHS (reproducing kernel Hilbert space) and implement PCA in that space. The resulting projections onto the kernel principal components can capture nonlinear patterns in the data. The KPCA procedure features fast training and fast scoring methods, which greatly alleviate the computation and memory burden associated with ordinary KPCA. These methods are based on low-rank matrix approximation coupled with k-means clustering as a sampling scheme. Thanks to this approximation algorithm, the KPCA procedure is capable of handling large data sets efficiently.

- The new SPARSEML procedure was designed for machine learning with sparse input data set. It currently implements the support vector machines (SVM) algorithm in SAS Viya for binary classification. By using the parallel coordinate descent optimization method, the SPARSEML procedure can train large data sets that can be both wide and deep. Like many other predictive modeling tools, the SPARSEML procedure uses input data to train a model and generates an analytic store that can be deployed through the ASTORE procedure. It can load data from multiple nodes and perform computations in parallel.
- The AUTOTUNE procedure now supports a hyperparameter importance ODS table, more options, saving the configuration history, and including warm start.
- The FACTMAC procedure now accepts data with missing values without issuing an error. Observations with missing values are excluded from the analysis.
- FOREST procedure
 - The NUMBIN= and MINLEAFSIZE= options can now be set by autotuning.
 - The default value of the NUMBIN= option is now 50 instead of 20, and the default value of the BINMETHOD= option is QUANTILE instead of BUCKET.
 - For isolation forests, the TARGET statement is no longer required, and fit statistics including variable importance are no longer computed.
- GRADBOOST procedure
 - The MINLEAFSIZE= option can now be set by autotuning.
 - The default value of some options have changed: the NUMBIN= option default is now 50 instead of 20; the MAXDEPT= option default is now 4 instead of 5; the RIDGE= option default is now 1 instead of 0; and the BINMETHOD= option default is now QUANTILE instead of BUCKET.
- The SEMISUPLEARN procedure now supports the AUTOTUNE statement.
- In the SVMACHINE procedure, the EARLYSTOP option is new. If the option is specified and the PARTITION statement is also specified, then the generated model is based on the validation accuracy.

SAS Visual Text Analytics 8.5

- Use stratified sampling when automatically generating categories for large data sets to reduce pipeline run time.
- Filter matches in the rule editor of the Edit Concept and Sandbox tabs by using the `_SELF_` keyword in conjunction with `REMOVE_ITEM` and `NO_BREAK` rules.
- Generate output data from a Concepts node and export it to SAS Visual Analytics for further analysis and visualization.
- Use a report template in SAS Visual Analytics to quickly create reports from Concepts and Categories output data.
- The autocomplete feature of the code editor in the interactive windows for the Concepts and Categories nodes is case sensitive, preventing accidental automatic completion of a key word as an operator.
- Add a pipeline template to a new or existing project in The Exchange.

For more information, see <http://documentation.sas.com/?activeCdc=ctxtcdc&cdclid=capcdc&cdcVersion=8.5&docsetId=ctxtug&docsetTarget=p0r4sjco27jrqn1khklf58sfidy.htm>

SAS Visual Text Analytics 8.5 includes these new programming features and enhancements:

- the ability to use the Sampling and Partitioning action set
- a redesigned rule generation action
- new action parameters in the Search action set, the ruleGen action, and the exportTextModel action
- the ability to search for special missing values
- improved capabilities for morphological expansion in the Korean language
- improvements to extraction of predefined concepts in the Danish language
- improved coverage of predefined concepts in the Spanish language

Link til alle analytics nyheder:

- https://go.documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=whatsnew&docsetTarget=titlepage.htm&locale=en

SAS Visual Forecasting 8.5

- The new DFIL procedure enables you to design a digital infinite impulse response (IIR) filter of any type, to visualize the frequency response of the designed filter, to visualize the zero-pole plot of the designed filter for stability analysis, and to filter any input data by using the designed digital filter.
- In the ARIMA statement of the SMSPEC procedure, the new P= and Q= options enable you to specify a custom range for the moving average (MA) and autoregressive (AR) orders of the ARIMA model specification.
- The new External Languages (EXTLANG) package enables you to integrate external-language code written in Python and R into your SAS program. The package supports various versions of the Python 2, Python 3, and R languages.
- The new Time Series Dimension Reduction (TDR) package enables you to reduce the dimensionality of time series. The dimension reduction technique includes piecewise aggregate approximation, symbolic aggregate approximation, discrete Fourier transformation, discrete wavelet transformation, random projection, and singular value decomposition.
- The Automatic Time Series Modeling (ATSM) package has been enhanced.
 - The new EventGroup method of the EVENT object enables you to specify a group event definition that is stored in the EVENT object.
 - The new 'CHOOSE' argument of the FORENG object enables you to specify the name of the winning forecast model.
 - The new 'ENDZEROS.MAXNUM', 'ENDZEROS.MAXPCT', and 'ENDZEROS.MINOBS' arguments of the FORENG object enable you to configure the _ZERO_ model test, which determines whether the FORENG object produces an all-zero forecast when trailing zeros are present in the dependent series.
 - The new 'ZEROMISS' options of the TSDF.AddSeries, TSDF.AddX, and TSDF.AddY methods enable you to specify how to interpret beginning and ending zero values in the ancillary, independent, and dependent series that you provide for the generation of forecast models

SAS Econometrics 8.5

- The new SASEMood interface engine enables SAS users to retrieve time series data from the Moody's Analytics Data Buffet (Economy.com) website.
- The new TSINFO procedure evaluates a variable in an input data table for its suitability as a time ID variable in SAS procedures and solutions that are used for time series analysis. PROC TSINFO assesses how well a time interval specification fits SAS date values, SAS datetime values, or observation numbers that are used to index a time series.

PROC TSINFO is intended for use as a tool to either identify the time interval of a variable or prepare problematic data sets for use in subsequent time series analyses. In particular, this procedure can be used to investigate inconsistencies between time ID values and the ID statement options that are used in other SAS procedures and solutions.

- The CCDM procedure has been enhanced. The output style as it relates to adjusted severity symbols is now consistent whether you specify one or more adjusted severity symbol in the ADJUSTEDSEVERITY= option. In previous releases, the output style with one adjusted severity symbol differed from the output style with more than one adjusted severity symbol. To revert to the older style for the case of one adjusted severity symbol, you can specify the new ONEADJSEVCOMPAT option in the PROC CCDM statement.
- CCOPULA procedure
 - The new PLOTS= option has been added to request correlation plots for pairs of variables in the DATA= data table for the FIT statement. Both the original data and their transformed uniform values can be plotted. The PLOTS= option is also available to plot the correlations between variables generated by the SIMULATE statement.
 - Additional options are available in the MARGINALS=EMPIRICAL option in the FIT statement. These options support a more accurate algorithm for computing the cumulative distribution function for variables in large data tables and for variables that have nonsmooth distributions and extreme values.
 - A new SimulateModelSummary ODS table has been added to describe results of the SIMULATE statement.
- The CPANEL procedure now supports these estimators:
 - heteroscedasticity-corrected standard error estimator (HCCME)
 - cross-sectional-correlation-corrected (CLUSTER) standard error estimator
 - heteroscedasticity- and autocorrelation-consistent (HAC) standard error estimator
- In the HMM procedure, state-independent parameter estimation and analytic store technology are now supported.
- In the SEVSELECT procedure, the parameter estimation phase now balances the available work in each BY group equitably among all available threads of computation. This new default behavior can lead to significantly faster estimation times, but it can also cause the numeric results to differ from those in previous releases. To revert to the numeric results of older releases, you can specify the new NOBALANCETHREADS option in the PROC SEVSELECT statement.