

SALFORD SYSTEMS

General Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Modeling Engine: CART (Decision Trees)	o	o	o	o
Modeling Engine: MARS (Nonlinear Regression)	o	o	o	o
Modeling Engine: TreeNet (Stochastic Gradient Boosting)	o	o	o	o
Modeling Engine: RandomForests for Classification	o	o	o	o
Reporting ROC curves during model building and model scoring	o	o	o	o
Model performance stats based on Cross Validation	o	o	o	o
Model performance stats based on out of bag data during bootstrapping	o	o	o	o
Reporting performance summaries on learn and test data partitions	o	o	o	o
Reporting Gains and Lift Charts during model building and model scoring	o	o	o	o
Automatic creation of Command Logs	o	o	o	o
Built-in support to create, edit, and execute command files	o	o	o	o
Translating models into SAS-compatible language (all other languages supported for Pro and above)	o	o	o	o
Reading and writing datasets in all current database/statistical file formats, including csv file format	o	o	o	o
Option to save processed datasets into all current database/statistical file formats	o	o	o	o
Select Cases in Score Setup	o	o	o	o
TreeNet Scoring Offset in Score Setup	o	o	o	o
Setting of focus class supported for all categorical variables	o	o	o	o
Scalable limits on terminal nodes. This is a special mode that will ensure the ATOM and/or MINCHILD	o	o	o	o
Descriptive Statistics: Summary Stats, Stratified Stats, Charts and Histograms		o	o	o
Activity Window: Brief data description, quick navigation to most common activities		o	o	o
Additional Modeling Engines: Regularized Regression (LASSO/Ridge/LARS/Elastic Net/GPS)		o	o	o
Data analysis Binning Engine		o	o	o

SALFORD SYSTEMS

General Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automatic creation of missing value indicators		o	o	o
Option to treat missing value in a categorical predictor as a new level		o	o	o
License to any level supported by RAM (currently 32MB to 1TB)		o	o	o
License for multi-core capabilities		o	o	o
Using built-in BASIC Programming Language during data preparation		o	o	o
Automatic creation of lag variables based on user specifications during data preparation		o	o	o
Automatic creation and reporting of key overall and stratified summary statistics for user supplied list of variables		o	o	o
Display charts, histograms, and scatter plots for user selected variables		o	o	o
Command Line GUI Assistant to simplify creating and editing command files		o	o	o
Translating models into SAS/PMML/C/Java/Classic and ability to create classic and specialized reports for existing models		o	o	o
Unsupervised Learning - Breiman's column scrambler		o	o	o
Scoring any Automate (pre-packaged scenario of runs) as an ensemble model		o	o	o
Summary statistics based on missing value imputation using scoring mechanism		o	o	o
Impute options in Score Setup		o	o	o
GUI support of SCORE PARTITIONS (GUI feature, SCORE PARTITIONS=YES)		o	o	o
Quick Impute Analysis Engine: One-step statistical and model based imputation		o	o	o
Advanced Imputation via Automate TARGET. Control over fill selection and new impute variable creation		o	o	o
Correlation computation of over 10 different types of correlation		o	o	o
Save OOB predictions from cross-validation models		o	o	o
Custom selection of a new predictors list from an existing variable importance report		o	o	o
User defined bins for Cross Validation		o	o	o

SALFORD SYSTEMS

General Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Cross-Validation models can now be scored as an Ensemble		o	o	o
An alternative to variable importance based on Leo Breiman's scrambler		o	o	o
Data Binning Results display (GUI feature)		o	o	o
Data Binning Analysis Engine bins variables using model-based binning (via AUTOMATE BIN), or using weights of evidence coding.		o	o	o
BIN ROUND, ADAPTIVEROUND methods (BIN METHOD=ROUND/ADAPTIVEROUND)		o	o	o
Controls for number of Bins and Deciles (BOPTIONS NBINS, NDECILES)		o	o	o
EVAL command and GUI display (GUI feature)		o	o	o
Summary stats for the correlations (Correlation Stats tab) (GUI feature)		o	o	o
TONUMERIC: create contiguous integer variables from other variables		o	o	o
Automation: Build two models reversing the roles of the learn and test samples (Automate FLIP)		o	o	o
Automation: Explore model stability by repeated random drawing of the learn sample from the original dataset (Automate DRAW)		o	o	o
Automation: For time series applications, build models based on sliding time window using a large array of user options (Automate DATASHIFT)		o	o	o
Automation: Explore mutual multivariate dependencies among available predictors (Automate TARGET)		o	o	o
Automated imputation of all missing values (via Automate Target)		o	o	o
Automation: Explore the effects of the learn sample size on the model performance (Automate LEARN CURVE)		o	o	o
Automation: Build a series of models by varying the random number seed (Automate SEED)		o	o	o
Automation: Explore the marginal contribution of each predictor to the existing model (Automate LOVO)		o	o	o
Automation: Explore model stability by repeated repartitioning of the data into learn, test, and possibly hold-out samples (Automate PARTITION)		o	o	o

SALFORD SYSTEMS

General Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automation: Explore the nonlinear univariate relationships between the target and each available predictor (Automate ONEOFF)		o	o	o
Automation: Bootstrapping process (sampling with replacement from the learn sample) with a large array of user options (Random Forests-style sampling of predictors, saving in-bag and out-of-bag scores, proximity matrix, and node dummies) (Automate BOOTSTRAP) *not available in RandomForests		o	o	o
Automation: AUTOMATE ENABLETIMING=YES NO to control timing reporting in Automates		o	o	o
Save out of bag predictions during Cross Validation			o	o
Use TREATMENT variables when scoring uplift models (SCORE EVAL)			o	o
Use TREATMENT variables when evaluating uplift model predictions (EVAL)			o	o
Automation: "Shifts" the "crossover point" between learn and test samples with each cycle of the Automate (Automate LTCROSSOVER)			o	o
Automation: Build a series of models using different backward variable selection strategies (Automate SHAVING)			o	o
Automation: Build a series of models using the forward-stepwise variable selection strategy (Automate STEPWISE)			o	o
Automation: Explore nonlinear univariate relationships between each available predictor and the target (Automate XONY)			o	o
Automation: Build a series of models using randomly sampled predictors (Automate KEEP)			o	o
Automation: Explore the impact of a potential replacement of a given predictor by another one (Automate SWAP)			o	o
Automation: Parametric bootstrap process (Automate PBOOT)			o	o
Automation: Build a series of models for each strata defined in the dataset (Automate STRATA)			o	o
Automation: Generate detailed univariate stats on every continuous predictor to spot potential outliers and problematic records (AUTOMATE OUTLIERS)			o	o
Automation: Convert (bin) all continuous variables into categorical (discrete) versions using a large array of user options (equal width, weights of evidence, Naïve Bayes, supervised) (AUTOMATE BIN)			o	o

SALFORD SYSTEMS

General Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automation: Build a series of models using every available data mining engine (Automate MODELS)				o
Automation: Run TreeNet for Predictor selection, Auto-bin predictors, then build a series of models using every available data mining engine (Automate GLM)				o
Modeling Pipelines: RuleLearner, ISLE				o

SALFORD SYSTEMS

Additional CART Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Modeling Engine: CART (Decision Trees)	o	o	o	o
Linear Combination Splits	o	o	o	o
Optimal tree selection based on area under ROC curve	o	o	o	o
User defined splits for the root node and its children		o	o	o
Translating models into Topology		o	o	o
Edit and modify the CART trees via FORCE command structures		o	o	o
RATIO of the improvements of the primary splitter and the first competitor		o	o	o
Scoring of CV models as an Ensemble		o	o	o
Report impact of penalties in root node		o	o	o
New penalty against biased splits PENALTY BIAS (PENALTY / BIAS, CONTBIAS, CATBIAS)		o	o	o
Automation: Generate models with alternative handling of missing values (Automate MISSING_PENALTY)		o	o	o
Automation: Build a model using each splitting rule (six for classification, two for regression) (Automate RULES)		o	o	o
Automation: Build a series of models varying the depth of the tree (Automate DEPTH)		o	o	o
Automation: Build a series of models changing the minimum required size on parent nodes (Automate ATOM)		o	o	o
Automation: Build a series of models changing the minimum required size on child nodes (Automate MINCHILD)		o	o	o
Automation: Explore accuracy versus speed trade-off due to potential sampling of records at each node in a tree (Automate SUBSAMPLE)		o	o	o
Automation: Generates a series of N unsupervised-learning models (Automate UNSUPERVISED)		o	o	o
Automation: Varies the RIN (Regression In the Node) parameter through the series of values (Automate RIN)		o	o	o
Automation: Varying the number of "folds" used in cross-validation (Automate CVFOLDS)		o	o	o
Automation: Repeat cross-validation process many times to explore the variance of estimates (Automate CVREPEATED)		o	o	o
Automation: Build a series of models using a user-supplied list of binning variables for cross-validation (Automate CVBIN)		o	o	o
Automation: Check the validity of model performance using Monte Carlo shuffling of the target (Automate TARGETSHUFFLE)		o	o	o

SALFORD SYSTEMS

Additional CART Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automation: Build two linked models, where the first one predicts the binary event while the second one predicts the amount (Automate RELATED). For example, predicting whether someone will buy and how much they will spend		o	o	o
Automation: Indicates whether a variable importance matrix report should be produced when possible (Automate VARIMP)		o	o	o
Automation: Saves the variable importance matrix to a comma-separated file (Automate VARIMPEFILE)		o	o	o
Automation: Generate models with alternative handling of missing values (AUTOMATE MVI)		o	o	o
Hotspot detection for Automate UNSUPERVISED			o	o
Hotspot detection for Automate TARGET			o	o
Hotspot detection to identify the richest nodes across the multiple trees			o	o
Differential Lift Modeling (Netlift/Uplift)			o	o
Profile tab in CART Summary window			o	o
Multiple user defined lists for linear combinations			o	o
Constrained trees			o	o
Ability to create and save dummy variables for every node in the tree during scoring			o	o
Report basic stats on any variable of user choice at every node in the tree			o	o
Comparison of learn vs. test performance at every node of every tree in the sequence			o	o
Automation: Vary the priors for the specified class (Automate PRIORS)			o	o
Automation: Build a series of models by progressively removing misclassified records thus increasing the robustness of trees and possibly reducing model complexity (Automate REFINE)			o	o
Automation: Bagging and ARcing using the legacy code (COMBINE)			o	o
Automation: Build a series of models limiting the number of nodes in a tree (Automate NODES)			o	o
Automation: Build a series of models trying each available predictor as the root node splitter (Automate ROOT)			o	o
Automation: Explore the impact of favoring equal sized child nodes by varying CART's end cut parameter (Automate POWER)			o	o
Automation: Explore the impact of penalty on categorical predictors (Automate PENALTY=HLC)			o	o

SALFORD SYSTEMS

Additional CART Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Build a Random Forests model utilizing the CART engine to gain alternative handling of missing values via surrogate splits (Automate BOOTSTRAP.BSPLIT)	o	o	o	o

SALFORD SYSTEMS

Additional MARS Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Modeling Engine: MARS (Nonlinear Regression)	o	o	o	o
Save MARS basis functions in Score Setup		o	o	o
MARS basis functions will be added during scoring to the output dataset.		o	o	o
Ridge parameter in MARS		o	o	o
Automation: Build a series of models varying the maximum number of basis functions (Automate BASIS)		o	o	o
Automation: Varying the number of "folds" used in cross-validation (Automate CVFOLDS)		o	o	o
Automation: Repeat cross-validation process many times to explore the variance of estimates (Automate CVREPEATED)		o	o	o
Automation: Build a series of models using a user-supplied list of binning variables for cross-validation (Automate CVBIN)		o	o	o
Automation: Build a series of models varying the smoothness parameter (Automate MINSPAN)			o	o
Automation: Build a series of models varying the order of interactions (Automate INTERACTIONS)			o	o
Automation: Build a series of models varying the modeling speed (Automate SPEED)			o	o
Automation: Explore the impact of penalty on categorical predictors (Automate PENALTY=HLC)			o	o
Automation: Explore the impact of penalty on missing values (Automate PENALTY=MISSING)			o	o
Automation: Build a series of models using varying degree of penalty on added variables (Automate PENALTY MARS)				o

SALFORD SYSTEMS

Additional TreeNet Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Modeling Engine: TreeNet (Stochastic Gradient Boosting)	o	o	o	o
Spline-based approximations to the TreeNet dependency plots		o	o	o
Exporting TreeNet dependency plots into XML file		o	o	o
Interactions: allow interactions penalty which inhibits TreeNet from introducing new variables (and thus interactions) within a branch of a tree		o	o	o
Auto creation of new spline-based approximation variables. One step creation and savings of transformed variable to new dataset		o	o	o
Automation: Build a series of models changing the minimum required size on child nodes (Automate MINCHILD)		o	o	o
Automation: Varying the number of "folds" used in cross-validation (Automate CVFOLDS)		o	o	o
Automation: Repeat cross-validation process many times to explore the variance of estimates (Automate CVREPEATED)		o	o	o
Automation: Build a series of models using a user-supplied list of binning variables for cross-validation (Automate CVBIN)		o	o	o
Automation: Check the validity of model performance using Monte Carlo shuffling of the target (Automate TARGET SHUFFLE)		o	o	o
Automation: Indicates whether a variable importance matrix report should be produced when possible (Automate VARIMP)		o	o	o
Automation: Saves the variable importance matrix to a comma-separated file (Automate VARIMPFIL)		o	o	o
Automation: Build a series of models by varying the random subsample size (Automate TNSUBSAMPLE)		o	o	o
Automation: Build a series of models by varying the quantile value when using the QUANTILE loss function(AutomateTNQUANTILE)		o	o	o
Automation: Build a series of models by varying the class weights between UNIT and BALANCED in N Steps (Automate TNCLASSWEIGHTS)		o	o	o
Auto creation of new spline-based approximation variables. One step creation and saving of transformed variable to new dataset.			o	o
Flexible control over interactions in a TreeNet model (ICL)			o	o
Interaction strength reporting			o	o
Interactions: Generate reports describing pairwise interactions of predictors			o	o
Interaction Control Lists (ICL): gives you complete control over structural interactions allowed or not allowed during model building.			o	o
Interactions: compute interaction statistics among predictors, for regression and logistic models only.			o	o
Subsample separately by target class. Specify separate sampling rates for the target classes in binary logistic models			o	o
Control number of top ranked models for which performance measures will be computed and saved			o	o

SALFORD SYSTEMS

Additional TreeNet Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Advanced controls to reduce required memory (RAM)			o	o
Extended Influence Trimming Controls: ability to limit influence trimming to focus class and/or correctly classified			o	o
Differential Lift Modeling (Netlift/Uplift)			o	o
Delta ROC Uplift as a performance measure			o	o
Uplift Profile tab for Uplift Results			o	o
TreeNet Newton Split Search and Regularization penalties (RGBBoost) (TN NEWTON=YES, RGLB0, RGLB1, RGLB2)			o	o
Save information for further processing and individual tree predictions			o	o
TreeNet Monotonicity Controls			o	o
Added Sample with Replacement option to GUI dialog			o	o
Hessian to control tree growing in TreeNet			o	o
Newton-style splitting is now available for TreeNet Uplift loss			o	o
QUANTILE specifies which quantile will be used during LOSS=LAD			o	o
POISSON: Designed for the regression modeling of integer COUNT data			o	o
GAMMA distribution loss, used strictly for positive targets			o	o
NEGBIN: Negative Binomial distribution loss, used for counted targets (0,1,2,3,...)			o	o
COX where the target (MODEL) variable is the non-negative survival time while the CENSOR variable indicates			o	o
Tweedie loss function			o	o
New table showing "Top Interactions Pairs"			o	o
Control over number of bins reported in Uplift tables			o	o
Translation of models with INIT option			o	o
Random Selection of Predictors: first for tree then random subset from that list for node			o	o
Save detailed 2-way interaction statistics to a file			o	o
Control the depth of each tree			o	o
Automation: Build two linked models, where the first one predicts the binary event while the second one predicts the amount (Automate RELATED). For example, predicting whether someone will buy and how much they will spend			o	o
Automation: Build a series of models limiting the number of nodes in a tree (Automate NODES)			o	o

SALFORD SYSTEMS

Additional TreeNet Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automation: Convert (bin) all continuous variables into categorical (discrete) versions using a large array of user options (equal width, weights of evidence, Naïve Bayes, supervised) (Automate BIN)			o	o
Automation: Produces a series of three TreeNet models, making use of the TREATMENT variable specified on the TreeNet command (Automate DIFFLIFT)			o	o
Automation: Build a series of models varying the speed of learning (Automate LEARNRATE)			o	o
Automation: Build a series of models by progressively imposing additivity on individual predictors (Automate ADDITIVE)			o	o
Automation: Build a series of models utilizing different regression loss functions (Automate TNREG)			o	o
Automation: Build a series of models by varying subsampling fraction (Automate TNSUBSAMPLE)			o	o
Automation: Build a series of models using varying degree of penalty on added variables (Automate ADDEDVAR)			o	o
Automation: Explore the impact of influence trimming (outlier removal) for logistic and classification models (AUTOMATE INFLUENCE)			o	o
Automation: Stochastic search for the optimal regularization penalties (Automate TNRGBOOST)			o	o
Modeling Pipelines: RuleLearner, ISLE				o
Build a CART tree utilizing the TreeNet engine to gain speed as well as alternative reporting, and control over interactions using ICL				o
Build a RandomForest model utilizing the TreeNet engine to gain speed as well as partial dependency plots, spline approximators, variable interaction statistics, and control over interactions using ICL				o
RandomForests inspired sampling of predictors at each node during model building				o
TreeNet Two-Variable dependency plots (3D plots) on-demand based on pairwise Interaction scores				o
TreeNet One-Variable dependency plots based on interaction scores				o
TreeNet in RandomForest mode for Classification				o
Random split selection (RSPLIT)				o
Median split selection (MSPLIT)				o

SALFORD SYSTEMS

Additional TreeNet Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automation: Explore the impact of influence trimming (outlier removal) for logistic and classification models (Automate INFLUENCE)				o
Automation: Exhaustive search and ranking for all interactions of the specified order (Automate ICL)				o
Automation: Varies the number of predictors that can participate in a TreeNet branch, using interaction controls to constrain interactions (Automate ICL NWAY)				o
Automation: Stochastic search of the core TreeNet modeling parameters (Automate TNOPTIMIZE)				o

SALFORD SYSTEMS

Additional Random Forests Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Modeling Engine: Random Forests for Classification	o	o	o	o
Additional Modeling Engine: Random Forests for Regression		o	o	o
Automation: Varies the bootstrap sample size (Automate RFBOOTSTRAP)		o	o	o
Automation: Vary the number of randomly selected predictors at the node-level (Automate RFNPREDs)		o	o	o
RF modified version of random split point selection (RANDOMMODE, JITTERSPLITS options)			o	o
Random Split Point is now exposed in GUI			o	o
Breiman's 2000 theory paper measures of STRENGTH and CORRELATION in the forest. (CORR, BCORR)			o	o
Penalty configuration for RF engine			o	o
RF: preserve prototype nucleus and consider variations to prototype algorithm (SVPROTOTYPES, PROTOREPORT)			o	o
GUI RF Advanced tab			o	o
in-bag / out-of-bag indicator to diagnostics dataset to facilitate testing (SVDIAG)			o	o
Reporting of "raw" permutation-based variable importance			o	o
Accuracy-based variable importance to RF, classification first			o	o
Saving of "margins" to output dataset (SVMARGIN)			o	o
Alternative, non-bootstrap forms of tree-by-tree sampling (SAMPLEAMOUNT, SAMPLEMODE, SAMPLEBYCLASS options)			o	o
New RF report: summarize N times each predictor appears in model, and N distinct split points			o	o
GUI controls for new Variable Importance measures			o	o
Flexible controls over interactions in a Random Forests for Regression model (requires TreeNet license)				o
Interaction strength reporting (requires TreeNet license)				o
Spline-based approximations to the Random Forests for Regression dependency plots (requires TreeNet license)				o
Exporting Random Forests for Regression dependency plots into XML files (requires TreeNet license)				o
Build a CART tree utilizing the Random Forests for Regression engine to gain speed as well as alternative reporting				o

SALFORD SYSTEMS

Additional Random Forests Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automation: Explore the impact of influence trimming (outlier removal) for logistic and classification models (Automate INFLUENCE)				o
Automation: Exhaustive search and ranking for all interactions of the specified order (Automate ICL)				o

SALFORD SYSTEMS

Additional Regression (OLS) Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Automation: Generate detailed univariate distributional reports for every continuous variable on the KEEP list (Automate OUTLIERS)			o	o

SALFORD SYSTEMS

Additional GPS Features

SPM 8.2 Feature Matrix

Components	Basic	Pro	ProEX	Ultra
Modeling Engines: Regularized Regression (LASSO/Ridge/LARS/Elastic Net/GPS)		o	o	o
Automation: Build a series of models by forcing different limit on the maximum correlation among predictors (Automate MAXCORR)			o	o