Microsoft Azure Machine Learning Release Notes

Microsoft Corporation

Published: July 16, 2014

This content is preliminary content. It might be incomplete and is subject to change.

Abstract

These notes encompass the Microsoft Azure Machine Learning Studio and Service (Public Preview), features, known issues and workarounds, future considerations, and contact information for feedback and support. Azure Machine Learning Studio is being offered for preview, which enables you to try out and provide feedback for 'in-concept' Azure tools, technologies, and feature sets.

Microsoft

Copyright Information

Contents

1 Summary

Microsoft Azure Machine Learning Studio (Public Preview) is a collaborative visual development environment that enables you to build, test, and deploy predictive analytics and machine learning solutions that operate on your data. The Azure Machine Learning is cloud-based, provides compute resource and memory flexibility, and offers zero-setup and installation woes because you work through your web browser.

The following components comprise Microsoft Azure Machine Learning:

- Rich user interface to create experiment and models.
- Large list of modules out of the box to build models.
- Extensibility by getting data from external sources and using R scripts for common analysis.
- The ability to prepare, clean, transform, and evaluate data against statistical, predictive, and social analytics models while using the power of the cloud from the convenience of your desktop.
- A mechanism for publishing workflows as staged or live request-response or batch processing evaluation or scoring service.

2 Features

Azure ML provides:

Feature	Description
Easy data import	Get data from a variety of sources including Windows Azure storage, SQL Azure, Hive, the Web, OData, and your own desktop.
Sample datasets and experiments for use 'right out of the box'	Sample datasets and experiments are provided that enable you to try out the machine learning algorithms.
Collaboration with others	Share your workspace with other users by invitation.
Extensibility with R	R-language script extensibility.
A library of modules to select from	Modules are applications that can read, or filter, or transform, or evaluate datasets that flow through your experiment. Azure ML Studio includes modules for social sentiment prediction, machine learning regressors, recommenders, and predictors, and extensibility (with the R statistical language).
Module documentation	Modules listed by function with input parameter descriptions.
Ability to stage and publish your models	You can publish and stage models as cloud-hosted batch or request response systems right from ML Studio.

Note!

For more information on the features, please consult the rest of the online help.

3 Providing Feedback and Getting Support

You can provide feedback or get support by using the link from the following table.

For this type of correspondence…	Use this …
Within the ML Studio UI.	Click the **Menu** button in the top right hand corner of the screen and choose **Feedback**. A free-form text input box displays. Tell us your experiences. Your feedback helps us prioritize improvements and features offered in the next refresh.
Azure Support	http://azure.microsoft.com/en-us/support/options/

4 Private Preview Customers: Migrating Existing Workflows to Public Preview Workspace

If you are a Private Preview Customer, and you have existing experiment workflows that you want to carry forward to your new Public Preview workspace, you can ask your TAP representative or the partner e-mail discussion list for instructions and help needed to migrate this information.

5 Known Issues and Workarounds

5.1 Storage key regeneration blocks workspace connectivity

When you provision an Azure ML workspace in the Azure Management Portal, a storage account and key are required. If you subsequently regenerate the key, the ML Studio users provisioned using that storage account will not be able upload datasets or run experiments.

There is no workaround at the user level. If you encounter this scenario, please contact CSS.

5.2 Stale autocomplete list in column selection utility after swapping datasets or modules

If you swap datasets or modules of your experiment workflow between subsequent runs, and these, the autocomplete dropdown list in column selection utility can show you columns present from the previous run.

Workarounds:

- Delete the module that is reducing the number (for example, Project Columns) and re-add. This clears assumptions of the columns configuration of the data at that point and following in the workflow.
- Re-run with the new modules and without reducing the columns (for example, for Project Columns this would mean by choosing “include all”).

5.3 ML Studio user workspace name (ID)

When renaming the WORKSPACE NAME field from the Settings tab within Azure ML Studio, to be consistent with Azure, these workspace names should not contain spaces only alpha numeric characters. Using spaces can break your subscription manager's ability to properly administer and manage your workspace.

Workarounds: Remove spaces for workspace name.

5.4 User is able to edit their profile in another user's shared workspace

After accepting an invitation, if the workspace invitation grants the invitee 'owner' status, these users can remove their invite profile from the other user's workspace.

Workaround: ask the user that shared his workspace to resend an invitation.

5.5 Selecting from more than 1000 columns

Performance in column selection from datasets with more than 1000 columns can be slow.

Workaround: Use rules for column selection; apply reduction steps outside of ML Studio.

5.6 Case sensitivity of Azure Storage Service account name

When provisioning workspaces in the Azure Management Portal, the workspace creation task expects a name that is the combination of lowercase and numbers. Using mixed case can result in failures.

Workaround: Create a new storage account with a name that is a mixture of lowercase and numbers.

5.7 (Resolved) Logging out from "Sign Out" link on the "Not Authorized" page

If you receive an error page that you do not have access to Azure ML because you are not logged in with a Microsoft user account (*@msn.com, *@outlook.com, etc.), the "Sign Out" link does not work properly.

Workaround:

- Do not click 'Sign out' on the page go to MSN.com, Outlook.com, or MSN.com and sign out.
- If you do click 'Sign out' delete your cookies from your browser data and then attempt to access Azure ML Studio or service.

5.8 Publishing is available for global samples, but errors out

If you try to publish a global sample experiment, you will receive an error. This is not a legal operation and the "Publish Web Service" button should be grayed out.

Workaround:

- Make a copy of the sample experiment - by clicking on "Save As" - then publish it as a Web Service.

5.9 Location is not visible in the Workspaces view in Azure

In the Workspaces view in Azure, the Location column is blank for some workspaces.

Workaround:

- There is currently no workaround for this. We are working on resolving this issue.

5.10 Intermittent error when running experiments

User may occasionally receive an error message “An error has occurred” when running an Experiment. This error goes away if the Experiment is re-run.

Workaround:

- Rerun the experiment.

5.11 RRS scoring fails with error 500 with no additional details on cause of the issue.

- Currently all failed scoring requests for a web service return a 500 error code with no additional details on the root cause of the issue.

  Workaround:

  When a 500 error is returned, the following steps should be performed to fix the issue:

  1. Make sure the input vector has exactly the same structure as defined in the web service API help page. Column names are case sensitive and all columns must be provided.
  2. When executing an experiment on a large amount of data, the service can make smarter decisions on the data types of each input column. When submitting a single vector for scoring, the service is doing best-effort type conversion. If during an experiment, a module was trained on a column of integers, and during the scoring, the feature was converted to double (which is the default for numeric columns), a module can fail.
     - To solve this issue, you should explicitly define the input types by adding a Metadata Editor module to your graph and explicitly converting your numeric data types to a specific data type.
  3. The Metadata Editor should be added before your scoring input port, and the input port selection should be changed to be the Metadata Editor input port. By doing so, you make sure that your input vector will always be converted to the correct data types.
  4. Example for modules which might fail due to a data type mismatch:
     - Score Model - If the model was trained on a dataset with integer types, the matching numeric features in the input vector should probably be converted to integers by using the Metadata Editor.
     - Modules with more than a single input port - If the module expects the same schema across multiple inputs and one of the input ports is getting its values from the input vector, you need to be sure that the input vector is converted to output with exactly the same schema.

6 Extensibility: Packages available to Execute R Module

For your convenience, the list of packages included in the current release is provided in the following table.

To get the complete list of packages that are currently available, see **Extensibility with R** in ML Studio help.

6.1 R packages beginning with A through E

Package name	Package description
abc	Tools for Approximate Bayesian Computation (ABC)
abind	Combine multi-dimensional arrays
actuar	Actuarial functions
ade4	Analysis of Ecological Data : Exploratory and Euclidean methods in Environmental sciences
AdMit	Adaptive Mixture of Student-t distributions
aod	Analysis of Overdispersed Data
ape	Analyses of Phylogenetics and Evolution
approximator	Bayesian prediction of complex computer codes
arm	Data Analysis Using Regression and Multilevel/Hierarchical Models
arules	Mining Association Rules and Frequent Itemsets
arulesViz	Visualizing Association Rules and Frequent Itemsets
ash	David Scott's ASH routines
assertthat	Easy pre and post assertions
AtelieR	A GTK GUI for teaching basic concepts in statistical inference, and doing elementary bayesian tests
BaBooN	Bayesian Bootstrap Predictive Mean Matching - Multiple and single imputation for discrete data
BACCO	Bayesian Analysis of Computer Code Output (BACCO)
BaM	Functions and datasets for books by Jeff Gill
bark	Bayesian Additive Regresssion Kernels
BAS	Bayesian Model Averaging using Bayesian Adaptive Sampling
base	The R Base Package
BayesDA	Functions and Datasets for the book *Bayesian Data Analysis*
bayesGARCH	Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations
bayesm	Bayesian Inference for Marketing/Micro-econometrics
bayesmix	Bayesian Mixture Models with JAGS
bayesQR	Bayesian quantile regression

Package name	Package description
bayesSurv	Bayesian Survival Regression with Flexible Error and Random Effects Distributions
Bayesthresh	Bayesian thresholds mixed-effects models for categorical data
BayesTree	Bayesian Methods for Tree Based Models
BayesValidate	BayesValidate Package
BayesX	R Utilities Accompanying the Software Package BayesX
BayHaz	R Functions for Bayesian Hazard Rate Estimation
bbemkr	Bayesian bandwidth estimation for multivariate kernel regression with Gaussian error
BCBCSF	Bias-corrected Bayesian Classification with Selected Features
BCE	Bayesian composition estimator: estimating sample (taxonomic) composition from biomarker data
bclust	Bayesian clustering using spike-and-slab hierarchical model, suitable for clustering high-dimensional data
bcp	A Package for Performing a Bayesian Analysis of Change Point Problems
BenfordTests	Statistical Tests for Evaluating Conformity to Benford's Law
bfp	Bayesian Fractional Polynomials
BH	Boost C++ header files
bisoreg	Bayesian Isotonic Regression with Bernstein Polynomials
bit	A class for vectors of 1-bit booleans
bitops	Bitwise Operations
BLR	Bayesian Linear Regression
BMA	Bayesian Model Averaging
Bmix	Bayesian Sampling for Stick-breaking Mixtures
BMS	Bayesian Model Averaging Library
bnlearn	Bayesian network structure learning, parameter learning and inference
boa	Bayesian Output Analysis Program (BOA) for MCMC
Bolstad	Bolstad functions
boot	Bootstrap Functions (originally by Angelo Canty for S)
bootstrap	Functions for the book *An Introduction to the Bootstrap*

Package name	Package description
bqtl	Bayesian QTL mapping toolkit
BradleyTerry2	Bradley-Terry Models
brew	Templating Framework for Report Generation
brglm	Bias reduction in binomial-response generalized linear models
bspec	Bayesian spectral inference
bspmma	bspmma: Bayesian Semiparametric Models for Meta-Analysis
BVS	Bayesian Variant Selection: Bayesian Model Uncertainty Techniques for Genetic Association Studies
cairoDevice	Cairo-based cross-platform antialiased graphics device driver
calibrator	Bayesian calibration of complex computer codes
car	Companion to Applied Regression
caret	Classification and Regression Training
catnet	Categorical Bayesian Network Inference
caTools	Tools: moving window statistics, GIF, Base64, ROC AUC, etc.
chron	Chronological objects which can handle dates and times
class	Functions for Classification
cluster	Cluster Analysis Extended Rousseeuw et al.
clusterSim	Searching for optimal clustering procedure for a data set
coda	Output analysis and diagnostics for MCMC
codetools	Code Analysis Tools for R
coin	Conditional Inference Procedures in a Permutation Test Framework
colorspace	Color Space Manipulation
combinat	combinatorics utilities
compiler	The R Compiler Package
corpcor	Efficient Estimation of Covariance and (Partial) Correlation
cslogistic	Conditionally Specified Logistic Regression
ctv	CRAN Task Views
cubature	Adaptive multivariate integration over hypercubes

Package name	Package description
data.table	Extension of data.frame
datasets	The R Datasets Package
date	Functions for handling dates
dclone	Data Cloning and MCMC Tools for Maximum Likelihood Methods
deal	Learning Bayesian Networks with Mixed Variables
Deducer	Deducer: A data analysis GUI for R
DeducerExtras	Additional dialogs and functions for Deducer
deldir	Delaunay Triangulation and Dirichlet (Voronoi) Tessellation.
DEoptimR	Differential Evolution Optimization in pure R
deSolve	General Solvers for Initial Value Problems of Ordinary Differential Equations (ODE), Partial Differential Equations (PDE), Differential Algebraic Equations (DAE), and Delay Differential Equations (DDE)
devtools	Tools to make developing R code easier
dichromat	Color Schemes for Dichromats
digest	Create cryptographic hash digests of R objects
distrom	Distributed Multinomial Regression
dlm	Bayesian and Likelihood Analysis of Dynamic Linear Models
doSNOW	Foreach parallel adaptor for the snow package
dplyr	dplyr: a grammar of data manipulation
DPpackage	Bayesian nonparametric modeling in R
dse	Dynamic Systems Estimation (time series package)
e1071	Misc Functions of the Department of Statistics (e1071), TU Wien
EbayesThresh	Empirical Bayes Thresholding and Related Methods
ebdbNet	Empirical Bayes Estimation of Dynamic Bayesian Networks
effects	Effect Displays for Linear, Generalized Linear, Multinomial-Logit, Proportional-Odds Logit Models and Mixed-Effects Models
emulator	Bayesian emulation of computer programs
ensembleBMA	Probabilistic Forecasting using Ensembles and Bayesian Model Averaging
entropy	Estimation of Entropy, Mutual Information and Related Quantities

Package name	Package description
EvalEst	Dynamic Systems Estimation - extensions
evaluate	Parsing and evaluation tools that provide more details than the default
evdbayes	Bayesian Analysis in Extreme Value Theory
evora	Epigenetic Variable Outliers for Risk prediction Analysis
exactLoglinTest	Monte Carlo Exact Tests for Log-linear models
expm	Matrix exponential
extremevalues	Univariate outlier detection

6.2 R packages beginning with F through L

Package name	Package description
factorQR	Bayesian quantile regression factor models
faoutlier	Influential case detection methods for factor analysis and SEM
fitdistrplus	Help to fit of a parametric distribution to non-censored or censored data
FME	A Flexible Modelling Environment for Inverse Modelling, Sensitivity, Identifiability, Monte Carlo Analysis
foreach	Foreach looping construct for R
forecast	Forecasting functions for time series and linear models
foreign	Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...
formatR	Format R Code Automatically
Formula	Extended Model Formulas
fracdiff	Fractionally differenced ARIMA aka ARFIMA(p,d,q) models
gam	Generalized Additive Models
gamlr	Gamma Lasso Regression
gbm	Generalized Boosted Regression Models
gclus	Clustering Graphics
gdata	Various R programming tools for data manipulation
gee	Generalized Estimation Equation solver
genetics	Population Genetics

Package name	Package description
geoR	Analysis of geostatistical data
geoRglm	geoRglm - a package for generalised linear spatial models
geosphere	Spherical Trigonometry
ggmcmc	Graphical tools for analyzing Markov Chain Monte Carlo simulations from Bayesian inference
ggplot2	An implementation of the Grammar of Graphics
glmmBUGS	Generalised Linear Mixed Models and Spatial Models with WinBUGS, BRugs, or OpenBUGS
glmnet	Lasso and elastic-net regularized generalized linear models
gmodels	Various R programming tools for model fitting
gmp	Multiple Precision Arithmetic
gnm	Generalized Nonlinear Models
googlePublicData	An R library to build Google's Public Data Explorer DSPL Metadata files
googleVis	Interface between R and Google Charts
GPArotation	GPA Factor Rotation
gplots	Various R programming tools for plotting data
graphics	The R Graphics Package
grDevices	The R Graphics Devices and Support for Colours and Fonts
gregmisc	Greg's Miscellaneous Functions
grid	The Grid Graphics Package
gridExtra	functions in Grid graphics
growcurves	Bayesian semi and nonparametric growth curve models that additionally include multiple membership random effects
grpreg	Regularization paths for regression models with grouped covariates
gsubfn	Utilities for strings and function arguments
gtable	Arrange grobs in tables
gtools	Various R programming tools
gWidgets	gWidgets API for building toolkit-independent, interactive GUIs
gWidgetsRGtk2	Toolkit implementation of gWidgets for RGtk2

Package name	Package description
haplo.stats	Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous
hbsae	Hierarchical Bayesian Small Area Estimation
hdrcde	Highest density regions and conditional density estimation
heavy	Package for outliers accommodation using heavy-tailed distributions
hflights	Flights that departed Houston in 2011
HH	Statistical Analysis and Data Display: Heiberger and Holland
HI	Simulation from distributions supported by nested hyperplanes
highr	Syntax highlighting for R
Hmisc	Harrell Miscellaneous
htmltools	Tools for HTML
httpuv	HTTP and WebSocket server library
httr	Tools for working with URLs and HTTP
IBrokers	R API to Interactive Brokers Trader Workstation
igraph	Network analysis and visualization
intervals	Tools for working with points and intervals
iplots	iPlots - interactive graphics for R
ipred	Improved Predictors
irr	Various Coefficients of Interrater Reliability and Agreement
iterators	Iterator construct for R
JavaGD	Java Graphics Device
JGR	JGR - Java GUI for R
kernlab	Kernel-based Machine Learning Lab
KernSmooth	Functions for kernel smoothing for Wand and Jones (1995)
KFKSDS	Kalman Filter, Smoother and Disturbance Smoother
kinship2	Pedigree functions
kknn	Weighted k-Nearest Neighbors
klaR	Classification and visualization

Package name	Package description
knitr	A general-purpose package for dynamic report generation in R
ks	Kernel smoothing
labeling	Axis Labeling
Lahman	Sean Lahman's Baseball Database
lars	Least Angle Regression, Lasso and Forward Stagewise
lattice	Lattice Graphics
latticeExtra	Extra Graphical Utilities Based on Lattice
lava	Linear Latent Variable Models
lavaan	Latent Variable Analysis
leaps	regression subset selection
LearnBayes	Functions for Learning Bayesian Inference
limSolve	Solving Linear Inverse Models
lme4	Linear mixed-effects models using Eigen and S4
lmm	Linear mixed models
lmPerm	Permutation tests for linear models
lmtest	Testing Linear Regression Models
locfit	Local Regression, Likelihood and Density Estimation
lpSolve	Interface to Lp_solve v. 5.5 to solve linear/integer programs

6.3 R packages beginning with M through R

Package name	Package description
magic	create and investigate magic squares
magrittr	magrittr - a forward-pipe operator for R
mapdata	Extra Map Databases
mapproj	Map Projections
maps	Draw Geographical Maps
maptools	Tools for reading and handling spatial objects

Package name	Package description
maptree	Mapping, pruning, and graphing tree models
markdown	Markdown rendering for R
MASS	Support Functions and Datasets for Venables and Ripley's MASS
MasterBayes	ML and MCMC Methods for Pedigree Reconstruction and Analysis
Matrix	Sparse and Dense Matrix Classes and Methods
matrixcalc	Collection of functions for matrix calculations
MatrixModels	Modelling with Sparse And Dense Matrices
maxent	Low-memory Multinomial Logistic Regression with Support for Text Classification
maxLik	Maximum Likelihood Estimation
mcmc	Markov Chain Monte Carlo
MCMCglmm	MCMC Generalised Linear Mixed Models
MCMCpack	Markov chain Monte Carlo (MCMC) Package
memoise	Memoise functions
methods	Formal Methods and Classes
mgcv	Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation
mice	Multivariate Imputation by Chained Equations
microbenchmark	Sub microsecond accurate timing functions
mime	Map filenames to MIME types
minpack.lm	R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds
minqa	Derivative-free optimization algorithms by quadratic approximation
misc3d	Miscellaneous 3D Plots
miscF	Miscellaneous Functions
miscTools	Miscellaneous Tools and Utilities
mixtools	Tools for analyzing finite mixture models
mlbench	Machine Learning Benchmark Problems
mlogitBMA	Bayesian Model Averaging for Multinomial Logit Models
mnormt	The multivariate normal and t distributions

Package name	Package description
MNP	R Package for Fitting the Multinomial Probit Model
modeltools	Tools and Classes for Statistical Models
mombf	Moment and Inverse Moment Bayes factors
monomvn	Estimation for multivariate normal and Student-t data with monotone missingness
monreg	Nonparametric monotone regression
mosaic	Project MOSAIC (mosaic-web.org) statistics and mathematics teaching utilities
MSBVAR	Markov-Switching, Bayesian, Vector Autoregression Models
msm	Multi-state Markov and hidden Markov models in continuous time
multcomp	Simultaneous Inference in General Parametric Models
multicool	Permutations of multisets in cool-lex order.
munsell	Munsell colour system
mvoutlier	Multivariate outlier detection based on robust methods
mvtnorm	Multivariate Normal and t Distributions
ncvreg	Regularization paths for SCAD- and MCP-penalized regression models
nlme	Linear and Nonlinear Mixed Effects Models
NLP	Natural Language Processing Infrastructure
nnet	Feed-forward Neural Networks and Multinomial Log-Linear Models
numbers	Number-theoretic Functions
numDeriv	Accurate Numerical Derivatives
openNLP	Apache OpenNLP Tools Interface
openNLPdata	Apache OpenNLP Jars and Basic English Language Models
OutlierDC	Outlier Detection using quantile regression for Censored Data
OutlierDM	Outlier detection for replicated high-throughput data
outliers	Tests for outliers
pacbpred	PAC-Bayesian Estimation and Prediction in Sparse Additive Models
parallel	Support for Parallel computation in R
partitions	Additive partitions of integers

Package name	Package description
party	A Laboratory for Recursive Partytioning
PAWL	Implementation of the PAWL algorithm
pbivnorm	Vectorized Bivariate Normal CDF
pcaPP	Robust PCA by Projection Pursuit
permute	Functions for generating restricted permutations of data
pls	Partial Least Squares and Principal Component regression
plyr	Tools for splitting, applying and combining data
png	Read and write PNG images
polynom	A collection of functions to implement a class for univariate polynomial manipulations
PottsUtils	Utility Functions of the Potts Models
predmixcor	Classification rule based on Bayesian mixture models with feature selection bias corrected
PresenceAbsence	Presence-Absence Model Evaluation
prodlim	Product-limit estimation. Kaplan-Meier and Aalen-Johansson method for censored event history (survival) analysis
profdpm	Profile Dirichlet Process Mixtures
profileModel	Tools for profiling inference functions for various model classes
proto	Prototype object-based programming
pscl	Political Science Computational Laboratory, Stanford University
psych	Procedures for Psychological, Psychometric, and Personality Research
quadprog	Functions to solve quadratic programming problems
quantreg	Quantile Regression
qvcalc	Quasi variances for factor effects in statistical models
R.matlab	Read and write of MAT files together with R-to-MATLAB connectivity
R.methodsS3	Utility function for defining S3 methods
R.oo	R object-oriented programming with or without references
R.utils	Various programming utilities
R2HTML	HTML exportation for R objects
R2jags	A Package for Running jags from R

Package name	Package description
R2OpenBUGS	Running OpenBUGS from R
R2WinBUGS	Running WinBUGS and OpenBUGS from R / S-PLUS
ramps	Bayesian Geostatistical Modeling with RAMPS
RandomFields	Simulation and Analysis of Random Fields
randomForest	Breiman and Cutler's random forests for classification and regression
RArcInfo	Functions to import data from Arc/Info V7.x binary coverages
raster	raster: Geographic data analysis and modeling
rbugs	Fusing R and OpenBugs and Beyond
RColorBrewer	ColorBrewer palettes
Rcpp	Seamless R and C++ Integration
RcppArmadillo	Rcpp integration for Armadillo templated linear algebra library
rcppbugs	R binding for cppbugs
RcppEigen	Rcpp integration for the Eigen templated linear algebra library
RcppExamples	Examples using Rcpp to interface R and C++
RCurl	General network (HTTP/FTP/...) client interface for R
relimp	Relative Contribution of Effects in a Regression Model
reshape	Flexibly reshape data
reshape2	Flexibly reshape data: a reboot of the reshape package
rgdal	Bindings for the Geospatial Data Abstraction Library
rgeos	Interface to Geometry Engine - Open Source (GEOS)
rgl	3D visualization device system (OpenGL)
RGraphics	Data and Functions from the book *R Graphics, Second Edition*
RGtk2	R bindings for Gtk 2.8.0 and above
RJaCGH	Reversible Jump MCMC for the analysis of CGH arrays
rjags	Bayesian graphical models using MCMC
rJava	Low-level R to Java interface
RJSONIO	Serialize R objects to JSON, JavaScript Object Notation

Package name	Package description
robCompositions	Robust Estimation for Compositional Data
robustbase	Basic Robust Statistics
RODBC	ODBC Database Access
rootSolve	Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations
roxygen	Literate Programming in R
roxygen2	In-source documentation for R
rpart	Recursive Partitioning and Regression Trees
rrcov	Scalable Robust Estimators with High Breakdown Point
rscproxy	statconn: provides portable C-style interface to R (StatConnector)
RSGHB	Functions for Hierarchical Bayesian Estimation: A Flexible Approach
RSNNS	Neural Networks in R using the Stuttgart Neural Network Simulator (SNNS)
RTextTools	Automatic Text Classification via Supervised Learning
RUnit	R Unit test framework
runjags	Interface utilities, parallel computing methods and additional distributions for MCMC models in JAGS
Runuran	R interface to the UNU.RAN random variate generators
rworldmap	Mapping global data, vector and raster
rworldxtra	Country boundaries at high resolution

6.4 R packages beginning with S through Z

Package name	Package description
SampleSizeMeans	Sample size calculations for normal means
SampleSizeProportions	Calculating sample size requirements when estimating the difference between two binomial proportions
sandwich	Robust Covariance Matrix Estimators
sbgcop	Semiparametric Bayesian Gaussian copula estimation and imputation
scales	Scale functions for graphics
scapeMCMC	MCMC Diagnostic Plots

Package name	Package description
scatterplot3d	3D Scatter Plot
sciplot	Scientific Graphing Functions for Factorial Designs
segmented	Segmented relationships in regression models with breakpoints/changepoints estimation
sem	Structural Equation Models
seriation	Infrastructure for seriation
setRNG	Set (Normal) Random Number Generator and Seed
sgeostat	An Object-oriented Framework for Geostatistical Modeling in S+
shapefiles	Read and Write ESRI Shapefiles
shiny	Web Application Framework for R
SimpleTable	Bayesian Inference and Sensitivity Analysis for Causal Effects from 2 x 2 and 2 x 2 x K Tables in the Presence of Unmeasured Confounding
slam	Sparse Lightweight Arrays and Matrices
smoothSurv	Survival Regression with Smoothed Error Distribution
sna	Tools for Social Network Analysis
snow	Simple Network of Workstations
SnowballC	Snowball stemmers based on the C libstemmer UTF-8 library
snowFT	Fault Tolerant Simple Network of Workstations
sp	classes and methods for spatial data
spacetime	classes and methods for spatio-temporal data
SparseM	Sparse Linear Algebra
spatial	Functions for Kriging and Point Pattern Analysis
spBayes	Univariate and Multivariate Spatial-temporal Modeling
spdep	Spatial dependence: weighting schemes, statistics and models
spikeslab	Prediction and variable selection using spike and slab regression
splancs	Spatial and Space-Time Point Pattern Analysis
splines	Regression Spline Functions and Classes
spTimer	Spatio-Temporal Bayesian Modelling Using R
stats	The R Stats Package

Package name	Package description
stats4	Statistical Functions using S4 Classes
stochvol	Efficient Bayesian Inference for Stochastic Volatility (SV) Models
stringr	Make it easier to work with strings
strucchange	Testing, Monitoring, and Dating Structural Changes
stsm	Structural Time Series Models
stsm.class	Class and Methods for Structural Time Series Models
SuppDists	Supplementary distributions
survival	Survival Analysis
svmpath	svmpath: the SVM Path algorithm
tau	Text Analysis Utilities
tcltk	Tcl/Tk Interface
tcltk2	Tcl/Tk Additions
TeachingDemos	Demonstrations for teaching and learning
tensorA	Advanced tensors arithmetic with named indices
testthat	Testthat code. Tools to make testing fun
textcat	N-Gram Based Text Categorization
textir	Inverse Regression for Text Analysis
tfplot	Time Frame User Utilities
tframe	Time Frame coding kernel
tgp	Bayesian treed Gaussian process models
TH.data	TH's Data Archive
timeDate	Rmetrics - Chronological and Calendar Objects
tm	Text Mining Package
tools	Tools for Package Development
translations	The R Translations Package
tree	Classification and regression trees
tseries	Time series analysis and computational finance

Package name	Package description
tsfa	Time Series Factor Analysis
tsoutliers	Automatic Detection of Outliers in Time Series
TSP	Traveling Salesperson Problem (TSP)
UsingR	Data sets for the text *Using R for Introductory Statistics*
utils	The R Utils Package
varSelectIP	Objective Bayes Model Selection
vcd	Visualizing Categorical Data
vegan	Community Ecology Package
VGAM	Vector Generalized Linear and Additive Models
VIF	VIF Regression: A Fast Regression Algorithm For Large Data
whisker	{{mustache}} for R, logicless templating
wordcloud	Word Clouds
XLConnect	Excel Connector for R
XML	Tools for parsing and generating XML within R and S-Plus
xtable	Export tables to LaTeX or HTML
xts	eXtensible Time Series
yaml	Methods to convert R data to YAML and back
zic	Bayesian Inference for Zero-Inflated Count Models
zipfR	Statistical models for word frequency distributions
zoo	S3 Infrastructure for Regular and Irregular Time Series (Z's ordered observations)

7 Additional Sample Datasets

The following additional sample datasets are used by sample experiments and are not listed in the ML Studio online help. The “Location column” qualifies them as either being available from the datasets palette in the ML Studio workspace or available from in Azure blob storage with links from the Machine Learning Center Models Gallery. For the other sample datasets included in Studio ML, see **Using Sample Datasets** in ML Studio online help.

Dataset Name	Sample Experiment	Location
News Categorization Train	News Categorization	Azure blob storage

News Categorization Test	News Categorization	Azure blob storage
News Categorization Labels	News Categorization	Azure blob storage
CRM Dataset	CRM	workspace
CRM Upselling Labels	CRM	workspace
CRM Churn Labels	CRM	workspace
CRM Appetency Labels	CRM	workspace
Network Intrusion Detection Dataset	Network Intrusion Detection	Azure blob storage
Flight Delays Data	Flight Delay Prediction	workspace
Airport Codes Dataset	Flight Delay Prediction	workspace
Weather Dataset	Flight Delay Prediction	workspace
German Credit Card UCI dataset	Credit Risk Prediction	workspace
German Credit Card UCI dataset	Credit Risk Prediction	Azure blob storage
Breast Cancer Features	Breast Cancer Detection	workspace
Breast Cancer Info	Breast Cancer Detection	workspace
Student Performance	Prediction of Student Performance	Azure blob storage
Direct Marketing Dataset	Direct Marketing via Uplift Modeling	Azure blob storage
Bike Rental UCI dataset	Prediction of the Number of Bike Rentals	workspace
Wikipedia SP 500 Dataset	Finding Similar Companies	workspace
Time Series Dataset	Time Series Prediction	workspace
Book Reviews from Amazon	Sentiment Classification	workspace
Bill Gates - RGB Image	Image Compression using Clustering	workspace
Movie Ratings	Movie Recommendation	workspace
Movie Tweets	Ranking of movie tweets according to their future popularity in Twitter	workspace
Movie Titles	Movie Recommendation, Ranking of movie tweets according to their future popularity in Twitter	workspace