

Machine Learning Cheatsheet

The tables below consist of valuable functions or commands that will help you through this module.

Each table represents a different library/tool and its corresponding commands.

Please note that these tables are not intended to tell you all the information you need to know about each command.

The hyperlinks found in each piece of code will take you to the documentation for further information on the usage of each command.

AnnotationDbi

Read the `AnnotationDbi` package vignette [here](#).

Library/Package	Piece of Code	What it's called	What it does
<code>AnnotationDbi</code>	<code>keytypes(.)</code>	Keytypes	Returns a character vector of column names/keytypes (eg., type of gene identifiers) available in an <code>AnnotationDbi</code> package.
<code>AnnotationDbi</code>	<code>mapIDs(.)</code>	Mapped IDs	Extracts the mapped ids for a set of keys (e.g., gene identifiers) of a specific keytype

Base

Read the Base  package documentation [here](#).

Library/Package	Piece of Code	What it's called	What it does
Base <code>R</code>	<code>round()</code>	Round	Rounds the values in the object provided in the first argument to the number of decimal places specified in the second argument
Base <code>R</code>	<code>identical()</code>	Identical	Checks if two objects are exactly equal
Base <code>R</code>	<code>prcomp()</code>	Principal Components Analysis	Executes a principal components analysis on specified matrix or data frame
Base <code>R</code>	<code>rowSums()</code>	Row Sums	Returns the sums of the rows in a numeric array, matrix, or data.frame
Base <code>R</code>	<code>rowMeans()</code>	Row Means	Returns the means of the rows in a numeric array, matrix, or data.frame
Base <code>R</code>	<code>quantile()</code>	Sample Quantiles	Returns the sample quantiles for a given numeric vector of data and numeric vector of probabilities
Base <code>R</code>	<code>cor()</code>	Correlation	Computes correlation between columns using a specified correlation method, and returns a correlation matrix
Base <code>R</code>	<code>as.dist()</code>	Distance matrix computation	Returns a special object of class <code>dist</code> , a distance matrix used by the <code>hclust()</code> function
Base <code>R</code>	<code>hclust()</code>	Hierarchical Clustering	Performs hierarchical clustering analysis on a set of dissimilarities and methods
Base <code>R</code>	<code>table()</code>	Create Table	Creates a contingency table of counts for each combination of factor levels
Base <code>R</code>	<code>duplicated()</code>	Duplicated	Returns a logical vector, where <code>TRUE</code> represents elements of the object that are duplicates
Base <code>R</code>	<code>any()</code>	Any	Checks to see if at least one of the elements are <code>TRUE</code> when given a logical vector
Base <code>R</code>	<code>cbind()</code>	Cbind	Combines vectors, matrices, or data.frames by columns
Base <code>R</code>	<code>pairwise.wilcox.test()</code>	Pairwise Wilcoxon Rank Sum Tests	Calculates the pairwise comparisons between group levels

PLIER

Read the `PLIER` package documentation [here](#).

A `PLIER` package vignette can be found [here](#) and can also serve as documentation for the commands in the table below.

Library/Package	Piece of Code	What it's called	What it does
<code>PLIER</code>	<code>combinePaths()</code>	Combine Pathways	Combines the pathway data obtained from <code>PLIER</code> and returns the result as a matrix
<code>PLIER</code>	<code>commonRows()</code>	Common Rows	Determines the rows (genes) that are common to the specified data matrices and returns them as a character vector
<code>PLIER</code>	<code>rowNorm()</code>	Row Normalize	Normalizes each row (gene) by z-scoring the expression values
<code>PLIER</code>	<code>num.pc()</code>	Number of Principal Components	Returns the number of significant principal components
<code>PLIER</code>	<code>PLIER()</code>	Main PLIER Function	Main function of the Pathway-Level Information ExtractoR.
<code>PLIER</code>	<code>plotU()</code>	Plot U Matrix	Plots the U matrix obtained from the <code>PLIER</code> function results, allowing insight into the pathways or cell types captured by the latent variables

ComplexHeatmap

Read the `ComplexHeatmap` package documentation [here](#).

Library/Package	Piece of Code	What it's called	What it does
<code>ComplexHeatmap</code>	Heatmap(.)	Complex Heatmap	Constructs a heatmap whose graphics and features can be defined
<code>ComplexHeatmap</code>	HeatmapAnnotation(.)	Heatmap Annotation Constructor	Creates an annotation object to be used in conjunction with a Heatmap

ggplot2

Read the `ggplot2` package documentation [here](#).

A vignette on the usage of the `ggplot2` package can be found [here](#).

Library/Package	Piece of Code	What it's called	What it does
<code>ggplot2</code>	geom_jitter(.)	Jittered Points	Adds a small amount of random variation at each point's location on a plot
<code>ggplot2</code>	labs(.)	Labels	Sets the axis, legend, and plot labels if specified
<code>ggplot2</code>	theme(.)	Theme	Sets the specified non-data elements of a plot (i.e. plot title, legend spacing, text size, etc.)

tidyr

Read the `tidyr` package documentation [here](#).

A vignette on the usage of the `tidyr` package can be found [here](#). | Library/Package | Piece of Code | What it's called | What it does | |-----|

`tidyr` | [separate\(.\)](#) | Separate | Separates a character column into multiple columns with a given regular expression or numeric locations || `tidyr` | [pivot_longer\(.\)](#) | Pivot Longer | Pivots data in a data.frame from wide to long format |

Other packages and functions

Documentation for each of these packages can be accessed by clicking the package name in the table below.

Library/Package	Piece of Code	What it's called	What it does
data.table	fread()	F read	Reads in data faster than base R
purrr	discard()	Discard	Discards the given elements
dplyr	pull()	Pull	Pulls a single variable out of a given table of data
matrixStats	rowSds()	Row Standard Deviations	Returns the standard deviation estimates for each row in a matrix
matrixStats	rowVars()	Row Variances	Returns the variance estimates for each row in a matrix
umap	umap()	Uniform Manifold Approximation and Projection (UMAP)	Computes a manifold approximation and projection on a given matrix or data.frame
ConsensusClusterPlus	ConsensusClusterPlus()	Consensus Clustering	Finds the consensus across multiple runs of the clustering algorithm
plotly	plot_ly()	Plotly Visualization	Initiates a plotly visualization with given R objects
ggsignif	geom_signif()	Create Significance Layer	Adds significance information to the plot. It can be used to run statistical tests and display the significance information from those tests. We use it differently, in a way that gives us more control, in the notebook.