

Intro to Machine Learning Cheat Sheet

Command	Description
!	Run bash command in Jupyter
import <package> [as <abbreviation>]	Import package
print()	Print out info
str()	Convert to string
int()	Convert to integer
float()	Convert to float
round()	Round number to desired number of digits
list()	Convert to list
Numpy: .array()	Create array
Numpy: .random.random()	Generate random number
Numpy: .setseed()	Set random seed so results will be the same
Numpy: .arange()	Create range of numbers
Pandas: DataFrame()	Create data frame
Pandas: read_csv()	Read csv into a data frame
Pandas: <dataframe>.to_csv()	Write dataframe to csv file
Pandas: <dataframe>.head()	Show first (5) rows of data frame
Pandas: <dataframe>.tail()	Show last (5) rows of data frame
Pandas: <dataframe>.sample()	Show sample row(s) of data frame
Pandas: <dataframe>.shape	Show number of rows and columns
Pandas: <dataframe>.columns	Show column names
Pandas: <dataframe>.dtypes	Show column data types
Pandas: <dataframe>.describe()	Show statistical summary of numerical cols

Pandas: <dataframe>.min/.max	Show min or max or column(s)
Pandas: <dataframe>.unique	Show unique values in column(s)
Pandas: <dataframe>.describe()	Show statistical summary of numerical cols
Pandas: <dataframe>[['colname']]	Index by column name(s)
Pandas: <dataframe>.iloc	Index rows (or columns) at particular positions(takes integers)
Pandas: <dataframe>.loc	Index rows (or columns) at particular positions or with labels
Pandas: factorize(<data>)	Encode data as categorical variable
Pandas: <dataframe>.replace()	Replace given input with given output throughout data frame
Pandas: <dataframe>.sort_values()	Sort values based on column name(s)
Pandas: <dataframe>.apply()	Apply a function to a column(s)
Pandas: <dataframe>.plot()	Plot data from a data frame
Pandas: <dataframe['column name']>.value_counts()	Returns counts of each unique value
Pandas: <dataframe>.rename()	Rename rows or columns in dataframe
Pandas: <dataframe>.copy()	Create copy of pandas object (use to prevent modifying a view)
Pandas: <dataframe>.merge()	Merges two data frames
Pandas: <dataframe>.isin()	Determine if element of data frame is contained in given values
Seaborn: boxplot(<data>)	Creates boxplot
Seaborn: pairplot(<data>)	Plot pairwise relationships in a dataset
Scikit-learn: <model>.fit(X,y)	Fits model with training data
Scikit-learn: <model>.predict(x)	Makes predictions based on trained model and test x values
Scikit-learn: <model>.score(y_true, y_pred)	Returns accuracy for classification model
Scikit-learn: <model>.predict_proba(x)	Makes probability predictions for classification

	model based on trained model and test x values
Scikit-learn: <code>metrics.precision_score(y_true, y_pred)</code>	Calculates precision
Scikit-learn: <code>metrics.recall_score(y_true, y_pred)</code>	Calculates recall
Scikit-learn: <code>metrics.accuracy_score(y_true, y_pred)</code>	Calculates accuracy
Pickle: <code>.dump(model)</code>	Serializes a python object (i.e. save a trained model to a file)
Pickle: <code>.load()</code>	Un-serializes a python object