

Summary of Utils.py Functions Relevant to the Introduction to Data Science Course

Function	Description	Related Topic
basic_stats(data)	Computes mean, std, skewness, and kurtosis of data.	Expectation and Moments
load_sms()	Loads the SMS spam dataset (text labeled ham/spam).	Pattern Recognition / Classification
discrete_histogram(data, normed=False)	Plots histogram of discrete values (optional normalization).	Exploratory Data Analysis
makeFreq(data_sequence)	Returns unique values with their frequencies.	Empirical Mass Function (EMF)
makeEMF(data_sequence)	Computes the empirical mass function (relative frequencies).	EMF and Discrete Distributions
plotEMF(numRelFreqPairs)	Plots the empirical mass function.	EMF Visualization
makeEDF(data_sequence)	Computes the empirical distribution function (cumulative frequencies).	Empirical Distribution Function (EDF)
emfToEdf(emf)	Converts EMF to EDF via cumulative sum.	EDF Construction
plotEDF(edf, confidence_band=False)	Plots EDF and optionally shows DKW confidence band.	Concentration Inequalities / Estimation
linConGen(m, a, b, x0, n)	Implements a linear congruential generator for pseudo-random numbers.	Simulation / Random Variable Generation
scatter3d(x, y, z)	Produces a 3D scatter plot using Plotly.	Visualization / Dimensionality Reduction
classification_report_interval(y_true, y_pred)	Computes precision, recall, accuracy with confidence intervals.	Pattern Recognition / Concentration
bennett_epsilon(n, b, sigma, alpha)	Computes deviation ϵ from Bennett's inequality.	Concentration Inequalities
epsilon_bounded(n, b, alpha)	Computes ϵ using Hoeffding's inequality for bounded variables.	Concentration Inequalities
compute_confidence_interval_bounded(data, delta, min, max)	Builds CI for the mean using Hoeffding bounds.	Estimation / Concentration
print_confidence_interval(point_estimate, epsilon)	Prints formatted confidence interval.	Estimation / Confidence Intervals
train_test_validation(X, Y)	Splits data into train, validation, and test sets.	Model Evaluation / ERM
timeout(func)	Decorator to limit execution time of a function.	Utility (Notebook Management)
showURL(url, ht=600)	Displays a web page or visualization in the notebook.	Visualization Utility