Code	**RAKING_ALGORITHM.PY**
Author	Nathaniel Heatwole, PhD (heatwolen@gmail.com) (GitHub) (LinkedIn)
Summary	Uses raking algorithm to shift (or benchmark) the output from a logistic regression equation (predicting student exam performance)
Methods/ Process	Raking algorithm - Iterative optimization routine for shifting relational row/column data, where both row- and column-level targets are known - Alternatively adjusts rows/columns, applying row- and column-specific multipliers, bidirectionally "raking" the data - As number of iterations increases, the values converge to the targets - Such an adjustment may be needed to address known or suspected bias in the model/data, or to align the predictions with other data - Used by government census agencies to assign person or household weights for national survey data (to account for non-response bias, and align with population-level totals from other data sources) Steps 1. Fits and assesses the (unmodified) predictions from the logistic regression[1] 2. *Column targets* (user-specified): total students predicted to pass exam (sum of predicted probabilities) 3. *Row targets*: total probability (pass/not pass) equals one 4. Apply column-specific multipliers (scalars) to hit column targets exactly (however, now row totals are misaligned with their targets) 5. Apply row-specific multipliers (scalars) to all variables to hit row targets exactly (however, now column totals are misaligned) 6. Repeat these steps, alternating rows/columns, either for some fixed number of iterations or until some convergence criteria are achieved
Training Data	Exam data – synthetic data for 20 students on whether they passed an exam and number of hours studied (from *Wikipedia*)
Results	Raking algorithm maintains the initial s-curve shape (logit) to the maximum extent, while simultaneously achieving all of the needed benchmarks/targets

[1] Owing to the logistic function's non-linear nature (s-curve), and its dual asymptotes at zero and one, the magnitude of this shift cannot be expressed in closed-form, and must be solved for iteratively.