Code	**RAKING_ALGORITHM.PY**
Author	Nathaniel Heatwole, PhD (heatwolen@gmail.com) (GitHub) (LinkedIn)
Summary	Uses a raking algorithm to shift/benchmark the output from a logistic regression equation (one 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 and columns, by applying row- and column-specific multipliers (adjustments), respectively (bidirectionally "raking" the data). - As the number of iterations increases, the values converge to the targets. - Such an adjustment may be needed for various reason, such as to address a known or suspected bias in the model/data. - Frequently used by government census agencies to assign person- or household-level weights for national survey data (to help account for non-response bias, and align various population-level totals with other data sources). Steps: 1. Import training data. 2. Assess the (unmodified) output from the logistic regression.[1] 3. *Column targets* (user-specified): number of students predicted to pass the exam (sum of predicted probabilities over all observations). 4. *Row targets*: total probability (pass, not pass) equals one. 5. Apply column-specific multipliers (scalars) to hit the column targets exactly. However, now the row totals are misaligned with their targets. 6. Apply row-specific multipliers (scalars) to all variables to hit the row targets exactly. However, now the column totals are misaligned with their targets. 7. Repeat these steps, alternating rows/columns, either for some fixed iterations or until some convergence criteria are achieved.
Training data	Exam data – synthetic data on whether 20 students passed an exam and number of hours studied (from Wikipedia).
Output	Plots: - Predicted probabilities (original, shifted) - Probability region added/removed Summary: - Predictions (original, shifted, target)
Result	The raking algorithm maintains, to the maximum extent, the initial s-curve shape, while simultaneously achieving all of the needed benchmarks/targets.

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