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 to address a known or suspected bias in the model or data, or to align the predictions with other 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 to align with various population-level totals from other data sources). Steps: 1. Assess the (unmodified) predictions from the logistic regression.[1] 2. *Column targets* (user-specified): number of students predicted to pass the exam (sum of predicted probabilities over all observations). 3. *Row targets*: total probability (pass, not pass) equals one. 4. Apply column-specific multipliers (scalars) to hit the column targets exactly. However, now the row totals are misaligned with their targets. 5. Apply row-specific multipliers (scalars) to all variables to hit the row targets exactly. However, now the column totals are misaligned. 6. Repeat these steps, alternating rows/columns, either for some fixed 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 that they studied (from *Wikipedia*).
Output	Plots: - Predicted probabilities (original, shifted) - Probability region added/removed Summary: - Predictions (original, shifted, target)
Result	The raking algorithm maintains the initial s-curve shape to the maximum extent possible, 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 amount cannot be expressed in closed-form, and must be solved for iteratively.