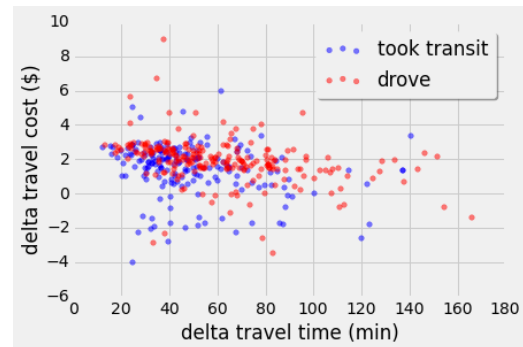


CE 88
Homework 7

In this homework, you will explore how data representation and scaling applied to the explanatory variables may affect the results produced by the predictive algorithms.

Data. We will continue working on the mode choice dataset explored in Minilab 7. Minilab 7 contains useful code for processing data and applying a simple k-nearest neighbors algorithm.



Problem 1 (5 points). Assume that the value-of-time (VOT) of all the travelers in the survey is \$20/hour. While it is often observed that VOT is proportional to income (and is approx. one half of the hourly pay rate), we will accept this common estimated value of \$20/hour as a reasonable simplifying assumption for this homework. Implement a data transformation function that scales travel time and costs into common units of either time or \$ (of your choice).

Problem 2 (5 points). Apply the k-nearest neighbor method to predict the travel mode using the VOT-scaled variables. Specifically predict the travel mode for a trip where the `delta_travel_time` = 50 minutes and `delta_tavel_cost` = 1.5.

- Report the predicted travel mode using the 20 nearest neighbors to make the prediction.
- Observe the indices of the 20 nearest neighbors, and report whether the nearest neighbors have changed from from the ones found when using the normalized inputs (as in mini-lab 7).

Your submission must contain both a PDF file describing your approach and the results, and the original ipynb.