CS506 HW0

1. Parse and Preprocess Data
   code in /src/1 - Import & Parse Data.py

2. Impute or Delete Missing Entries
   (a) def impute_missing(X):
   code in /src/2 - Impute or Delete Missing Entries.py

   (b) Mean is particularly susceptible to the influence of outliers. When there are values that are unusual compared to the rest of the data set, it is better to use the median instead of the mean for missing values.

   (c) def discard_missing(X, y):
   code in /src/2 - Impute or Delete Missing Entries.py

3. Working with the Data
   Assume the input to all functions is represented as a matrix – X.
   (a) def shuffle_order(X):
   code in /src/3 - Working with the Data.py

   (b) def stdev(X):
   code in /src/3 - Working with the Data.py
   Note: the function returns a list containing the standard deviation of each feature.

   (c) def remove_more_than_two_std(X):
   code in /src/3 - Working with the Data.py

   (d) def standardize(X):
   code in /src/3 - Working with the Data.py
   Note: the function reuses the function stdev in (b)
   The time complexity for this function is $O(n^2)$ and the space complexity is $O(n^2)$.

4. Working with Non-numerical Data
   def import_non_num_data(dataset):
   code in /src/4 - Working with Non-numerical Data.py

5. Train-Test Split
   (a) def train_test_split(X, y, t_f):
   code in /src/5 - Dataset Split.py

   (b) def train_test_CV_split(X, y, t_f, cv_f):
   code in /src/5 - Dataset Split.py