Week 3 Project: Implementation FAQs

- **How to read the files ? Can we use pandas ?**

  Note the code to read the files are already given.

  X_train = np.genfromtxt(sys.argv[3], delimiter = ",")

  y_train = np.genfromtxt(sys.argv[4])

  X_test = np.genfromtxt(sys.argv[5], delimiter = ",")

  Please use this code, to appropriately read the data. This will ensure that each row of X_train and X_train is a vector and y_train's each row is a scalar float.

- **How to save the output ?**

  See the following code -

  wRR = part1() # Assuming wRR is returned from the function

  np.savetxt("wRR_" + str(lambda_input) + ".csv", wRR, delimiter="\n") # write output to file

  active = part2() # Assuming active is returned from the function

  np.savetxt("active_" + str(lambda_input) + "_" + str(int(sigma2_input)) + ".csv", active, delimiter=",") # write output to file

  It already prints the data in a proper format.

- **What are we supposed to output for task 2 ?**

  Index of X vectors where you will the ground truth. Ie points of maximum uncertainty.

- **What floating point precision levels to use ?**

  Use standard floats for weight vector. For locations (part 2) you can use integer.

- **What Python version to use ?**

  Only use python 3.x.

- **What libraries we can use ?**

  For this problem only numpy is needed. Other non standard libraries might not work.

- **Where to get test data sets ?**

  You can get various regression data sets from here : http://archive.ics.uci.edu/ml/

- **Can we submit any number of time ?**

  You can submit any number of times.

- **What does this error mean ?**

  ## standardization formula ##

  error: dlmread: unable to open file 'active_8_3.csv'

  In general this errors come from the grader. But this does not mean the file is unreadable. Normally the file is not created at all, as your code might have an early exit due to some exception. You should do exception handling to see what exactly is the error.

- **What exactly to do in Part 2 ?**

  Compute the covariance matrix. Select the location which maximizes uncertainty. Update the covariance matrix. Among the remaining points, select the index of the one which again maximizes uncertainty. Update covariance. And so on.