CS5691: Pattern recognition and machine learning
Programming Assignment 2
Course Instructor : Arun Rajkumar.
Release Date : October-09, 2019
Submission Date: On or before 5 PM on October-27,2019

SCORING: There are 3 questions in this assignment. Each question carries 5 points. The total points obtained will be multiplied by $\frac{2}{3}$ as contribution towards your final grades. The points will be decided based on the report provided, code submitted and a final oral examination that covers all the assignments together.

DATASETS Check the README file for description.

WHAT SHOULD YOU SUBMIT? You should submit a zip file titled 'Solutions_ rollnumber1 _rollnumber2.zip' where rollnumber1 and rollnumber2 are roll numbers of the members of the group. Your assignment will NOT be graded if it does not contain all of the following:

- A text file titled 'Participants.txt' with names and roll numbers of members.
- A PDF file which includes explanations regarding each of the solution as required in the question. Title this file as 'Report.pdf'
- Source code for all the programs that you write for the assignment clearly named.

CODE LIBRARY: You are expected to code all algorithms from scratch. You cannot use standard inbuilt libraries for **computations**. The only allowed library are those that compute the Eigenvectors and Eigenvalues of matrices. If your code calls any other library function for computation, it will fetch 0 points. You are free to use inbuilt libraries for plots. You can code using either Python or Matlab or C.

GUIDELINES: Keep the below points in mind before submission.

- Plagiarism of any kind is unacceptable. These include copying text or code from any online sources. These will lead to disciplinary actions according to institute guidelines.
- Any graph that you plot is unacceptable for grading unless it labels the x-axis and y-axis clearly.
- Don't be vague in your explanations. The clearer your answer is, the more chance it will be scored higher.

LATE SUBMISSION POLICY You are expected to submit your assignment on or before the deadline to avoid any penalty. Late submission incurs a penalty equal to the number of days your submission is late by. For instance if you score 12 points out of 15, then your non-penalized score out of 10 would be 8. However, If you submit it after 5 PM on Oct-27,2019 and before 5 PM on Oct-28, 2019, your score will be $8 - 1 = 7$ points. If you submit it after 5 PM on Oct-28 and before 5 PM on Oct 29, your score will be $8 - 2 = 6$ points and so on. If you obtain negative points after penalty (or don't turn in your assignment), it will be considered as 0 points.

QUESTIONS

(1) You are given a data-set with 1000 data points each in $\mathbb{R}^2$.

i. Write a piece of code to run the Llyod's algorithm for the K-means problem with $k = 4$. Try 5 different random initialization and plot the error function w.r.t iterations in each case. In each case, plot the clusters obtained in different colors.

ii. Fix a random initialization. For $K = \{2, 3, 4, 5\}$, obtain cluster centers according to Lloyd's algorithm using the fixed initialization. For each value of K, plot the Voronoi regions associated to each cluster center. (You can assume the minimum and maximum value in the data-set to be the range for each component of $\mathbb{R}^2$).

iii. Run the spectral clustering algorithm (spectral relaxation of K-means using Kernel-PCA) $k = 4$. Choose an appropriate kernel for this data-set and plot the clusters obtained in different colors. Explain your choice of kernel based on the output you obtain.

(2) You are given a data-set with 10000 points in $(\mathbb{R}^{100}, \mathbb{R})$ (Each row corresponds to a datapoint where the first 100 components are features and the last component is the associated y value).

i. Obtain the least squares solution $\mathbf{w}_{ML}$ to the regression problem using the closed form expression.

ii. Code the gradient descent algorithm with suitable step size to solve the least squares algorithms and plot $\|\mathbf{w}^t - \mathbf{w}_{ML}\|$ as a function of t. What do you observe?

iii. Code the stochastic gradient descent algorithm using batch size of 100 and plot $\|\mathbf{w}^t - \mathbf{w}_{ML}\|$ as a function of t. What are your observations?

(3) Consider the same data-set as in Question (2). You are additionally given a data-set with 500 points for testing which you cannot use during train/cross-validation.

i. Code the gradient descent algorithm for ridge regression.

ii. Cross-validate for various choices of λ and plot the error in the validation set as a function of λ. For the best λ chosen, obtain $\mathbf{w}_R$. Also obtain $\mathbf{w}_{ML}$ for the training data. Compare the test error of $\mathbf{w}_R$ with $\mathbf{w}_{ML}$. Which is better and why?

iii. Code the co-ordinate descent algorithm to obtain a LASSO solution. Cross validate for the same train-validation splits as in part (ii). Plot the error in the validation set as a function of λ and obtain the best λ. Compare the test error of $\mathbf{w}_{LASSO}$ with $\mathbf{w}_R$. Which is better and why?