Programming Assignment # 2

Due: Thursday, March 9, 2023, 11:59 p.m.
Total Points: 100

Grader: Junyi Fan
Office Hours: TW 3:30pm - 4:30pm (Zoom)
Contact: Slack

This assignment is designed to give you practical programming experience with radial basis functions and the K-means algorithm. Many aspects of this class can be fully understood only with computer simulations. Programming assignments help you gain first-hand experience with the algorithms introduced in the class. Please use Python for the implementation, but you CANNOT use any library that implements the solution for you (e.g., for K-Means or RBF). Message the TA/instructor in Slack if you are unsure. If you are unfamiliar with Python, please reach out to the instructor and grader as soon as possible.

Please carefully read all instructions below and also periodically check Slack for updates. Any questions that you post to the public Slack channels should NOT contain any portion of your solution. If you need specific help with your solution, you can send private messages to the instructor or grader. Post general questions to the programming-2 Slack channel.

This is an individual assignment, so you must submit your own work. You also cannot use other resources (e.g., web sites, books, blogs, research papers, other people, etc.) to solve the problems. **Your assignment must be submitted by committing your code to your Github repository before the deadline. Please create a folder called 'pa2' (without quotes) and commit all files for this assignment to this folder.** You are strongly encouraged to submit early and often to avoid any last minute issues and to confirm that you are committing your work correctly. The submission should be in the form of a Python 3 Jupyter Notebook file. **Be sure to run the file before committing, so that we can directly see your results**. Programs that were not run before submitting will result in a zero.

This assignment must be submitted on time to receive full credit. Late submissions will result in a 10% penalty each calendar day it is late, but it cannot be submitted more than 3 days late. **Your grade will partially depend on completeness, correctness, code efficiency (e.g., using linear algebra whenever possible), and code readability (e.g., comments).**

Question 1. [100 POINTS]

Implement an RBF network for a one-dimensional input variable, one-dimensional output variable, and Gaussian basis functions. Generate a set of 75 data points by sampling the function $h(x) = 0.5 + 0.4sin(2\pi x)$ with added uniform noise in the interval [-0.1, 0.1] and with x values taken randomly from a uniform distribution in the interval [0.0, 1.0]. Determine the Gaussian centers by the K-means algorithm, and set the variance of each cluster accordingly. If a cluster contains only one sample point, use as its variance the mean variance of all the other clusters. Use the LMS rule for weight update (note a bias term is needed). Perform the following:

- Vary the number of bases in the range of 2, 4, 7, 11, and 16.
- Use two values of η: 0.01 and 0.02.

For each of the above 10 cases, stop training after 100 epochs. Plot for each case a graph that shows the data points, the original function where the data points are sampled from, and the function generated by the RBF network.

Comment on the dependence of the network performance on the number of bases. Also comment on the choice of η.

In addition, repeat the above implementation but use the same variance for all clusters as described in the lectures. Comment on the comparative advantages of the two methods of determining cluster variance.

What you need to turn in:

1. A summary report that is at least two-pages long, which contains your test results, along with a description and explanation of those results.
2. Your Jupyter notebook that contains your code and results.