CS 240: Lab 8
Kernelized Linear, Logistic and Softmax Regression

TAs: Jaladhi Joshi, Thomas Biju

Instructions

- This lab will be **graded**.
- Please read the problem statement and submission guidelines carefully.
- For any doubts or questions, please contact the assigned TA.
- The deadline for this lab is **Thursday, 20 March, 5 PM**. Solutions till 5:30 PM will be accepted. No submissions will be accepted after 5:30 PM.
- The submissions will be checked for plagiarism, and any form of cheating will be penalized.

Problem Statement

Part 1: Kernelized Linear Regression

The first task is to implement kernelized linear regression on a given synthetically generated dataset.

- The dataset contains 200 training samples, 50 validation samples, with each input and output being one-dimensional, following a polynomial function of degree 5: $y = f(x) = \sum_{i=0}^{5} a_i x^i$
- Define a degree 5 polynomial kernel $\phi(x)$ mapping from one degree input to 5 degree feature vector and using this as the input, run Linear Regression to find the polynomial coefficients.
- Plot the data as well as the outputs predicted by the model, for visualization. Also, print the MSE among the validation samples.
- You are allowed to use the `sklearn` **Linear Regression** package.

Part 2: Logistic and Softmax Regression on MNIST

1. **Logistic Regression:** Perform **one-vs-all** classification on the MNIST dataset.
   - Choose one class as positive and the rest as negative, and train a logistic regression model for binary classification. Do this step for each class and compute the accuracies.
   - You are allowed to use `sklearn` Logistic Regression package for this task.
2. **Softmax Regression:** Implement softmax regression on the MNIST dataset.
   - Complete the required functions, following the comments, in the provided template file.
   - You are not allowed to use external regression Python packages for this part.

Compare the accuracies of logistic regression and softmax regression for each class.

Tasks to be Completed

Complete the following tasks in the provided Jupyter Notebook file:

Task 1 - Implement kernelized linear regression on the synthetic dataset, by using the polynomial kernel of degree 5. Plot the datapoints as well as the predicted outputs on the given inputs. Also, print the MSE on the validation dataset.

Task 2 - Perform one-vs-all logistic regression on MNIST and report accuracy for each class. You are allowed to use `sklearn` Logistic Regression package.

Task 3 - Implement softmax regression on MNIST, by completing the given template file, and report the accuracy for each class. Compare these accuracies with those obtained by logistic regression.

Submission

- Submissions should be made on Moodle. Submit the Jupyter Notebook file renamed as `rollnumber1_rollnumber2.ipynb` (the "b" in roll number should be in small case).
- The hard deadline for submission is 5:30 pm. No submission after that will be evaluated.
- Only one person per team should submit their solution.