Assignment 4

September 2020

Instructions

- This assignment should be completed individually.
- Do not look at solutions to this assignment or related ones on the Internet.
- The files related to the assignment are present in `lab4-rollno.zip` folder. Extract it and upload it on moodle in the same .zip format after completion and after replacing the string "rollno" with your actual roll number. For example, if you roll number is 00405036, then single zip folder that you will upload will be named "lab4-00405036.zip". Also collate all the CS337 based theory solutions into ONE pdf file named `answers.pdf`. Include `answers.pdf` inside the zip folder mentioned above and submit the zip folder.
- Answers to all subjective questions need to be placed in single pdf `answers.pdf` including all plots and figures and uploaded.
- Only add/modify code between `TODO` and `END TODO` unless specified otherwise. You must not import any additional libraries.
- Python files to submit - `kernel.py`, `kernel_logistic.py`, `krr.py` and `kmeans.py`
- This Assignment carries a total of **8** marks for CS337 Theory and **14.5** marks for CS335 Lab

1 Kernel Methods

1.1 CS 337: Theoretical Problem

Prove that the function $K_\sigma : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined as $K_\sigma(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(\frac{-\|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\sigma^2}\right)$ is a valid Kernel. You may use the properties proved in class. [Hint - Taylor series expansion] (2 marks)

1.2 CS 335: Lab Problems

In this task, you need to complete functions in the files `kernel.py`, `krr.py` and `kernel_logistic.py`. The details for each subtask are listed below-

(a) **kernel.py** - Complete the functions `gaussian_kernel` and `linear_kernel` to return the Gram matrix. For vectors x and y in $\mathbb{R}^d$, the linear kernel is given by $K(x, y) = \sum_{i=1}^{d} x_i y_i$ and the gaussian kernel is given by $K_\sigma(x, y) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right)$ (2 marks)

(b) **kernel_logistic.py** -

(i) Complete the functions `fit` and `predict` to train a logistic regression using the gaussian kernel. Use gradient descent to minimize the dual kernelized objective of logistic regression as presented in Lecture Lecture 10 . (1.5 marks)

(ii) Complete the function `k_fold_cv` to perform k-fold cross validation to get the best σ. Divide the dataset into k equal parts for this without any randomization. Report the best σ obtained along with the plot in the report. Ensure that you use the default values of η, λ, max_iter as provided in the starter code. (1.5 marks)

(iii) Briefly justify the plot of the variation of error and sigma that you obtained. (1 mark)

(c) **krr.py** -

(i) Complete the functions `fit` and `predict`. These need to be done without any for loops. (1.5 marks)

(ii) Put the first two plots obtained depicting variation of fit with σ and λ in the report. Interpret the graph produced by various values of λ and σ. (1.5 marks)

2 Kernel Design

2.1 CS 337: Theory Questions

Given that $K(\mathbf{x}, \mathbf{x}')$ is a valid kernel, where $\mathbf{x} \in \mathcal{R}^m$, prove the following -

(i) Let $g : \mathcal{R}^m \to \mathcal{R}^m$ be a function. Then show that $K(g(\mathbf{x}), g(\mathbf{x}'))$ is a valid kernel.(1.5 marks)

(ii) Let q be a polynomial with non-negative coefficients. Then show that $q(K(\mathbf{x}, \mathbf{x}'))$ is a valid kernel. (1.5 marks)

2.2 CS 335: Lab Question

In this task, you need to design a kernel to fit the given data. Complete `my_kernel()` function in `kernel.py` such that a good fit to the data is obtained in the third figure plotted when running `krr.py`. Mention your kernel function in the report along with the obtained plot. To get an idea of what type of kernel to use, you may plot variation of the target value of the data with y at a constant x or vice versa. (1.5 marks)

3 K-means clustering for Image compression

3.1 CS 337: Theory Questions

Consider an optimal 2-clustering of a data set $x^1, x^2, \ldots, x^n$, $n \geq 2$, where for $i \in \{1, 2, \ldots, n\}$, $x^i \in \mathbb{R}^d$, where $d \geq 1$. Without loss of generality, assume that for some $m \in \{1, 2, \ldots, n-1\}$, the points

$x^1, x^2, \ldots, x^m$ are assigned to the first cluster, and the points $x^{m+1}, x^{m+2}, \ldots, x^n$ are assigned to the second cluster. Assume that the n points are all distinct, and no point is equidistant to both cluster centres.

Show that there exists a hyperplane of the form $a \cdot x + b = 0$, where $a \in \mathbb{R}^d, b \in \mathbb{R}$, such that all the points in the first cluster $(x^1, x^2, \ldots, x^m)$ lie on one side of the hyperplane and all the points in the second cluster $(x^{m+1}, x^{m+2}, \ldots, x^n)$ lie to the other side of the hyperplane. You should express a and b in terms of $x^1, x^2, \ldots, x^n$, n and m. (3 marks)

3.2 CS 335: Lab Question

(i) In `kmeans.py`, complete the functions `init_clusters`, `pred` and `train` (2 marks)

(ii) For each of the given 3 images, comment on the relation between number of clusters and quality of the generated images, *i.e.* generate images for $k = 2, 5, 10$ for all three images and comment on the resultant image. Attach all 9 images in the report. (1 marks)

(iii) Describe why some generated images look fine with smaller number of clusters while others need more clusters (1 marks)