IITM-CS5691 : Pattern Recognition and Machine Learning — Release Date: October 9, 2023
Assignment II — Due Date : October 23, 2023, 23:59

Roll No: CS23E001 — Name: Shuvrajeet Das
Collaborators (if any):
References/sources (if any):

- Use LaTeX to write-up your solutions (in the solution blocks of the source LaTeX file of this assignment), submit the resulting rollno.asst2.answers.pdf file at Crowdmark by the due date, and propery drag that pdf's answer pages to the corresponding question in Crowdmark (do this propery, otherwise we won't be able to grade!). (Note: **No late submissions** will be allowed, other than one-day late submission with 10% penalty or four-day late submission with 30% penalty.)
- Please upload to moodle a rollno.zip file containing three files: rollno.asst2.answers.pdf file mentioned above, and two code files for the programming question (rollno.ipynb file and rollno.py file). Do not forget to upload to Crowdmark your results/answers (including Jupyter notebook **with output**) for the programming question.
- Collaboration is encouraged, but all write-ups must be done individually and independently, and mention your collaborator(s) if any. Same rules apply for codes written for any programming assignments (i.e., write your own code; we will run plagiarism checks on codes).
- If you have referred a book or any other online material or LLMs (Large Language Models like ChatGPT) for obtaining a solution, please cite the source. Again don't copy the source *as is* - you may use the source to understand the solution, but write-up the solution in your own words (this also means that you cannot copy-paste the solution from LLMs!). Please be advised that *the lesser your reliance on online materials or LLMs* for answering the questions, *the more your understanding* of the concepts will be and *the more prepared you will be for the course exams*.
- Points will be awarded based on how clear, concise and rigorous your solutions are, and how correct your answer is. The weightage of this assignment is 12% towards the overall course grade.

1. (8 points) [LINEAR REGRESSION]

   (a) (4 points) The error function in the case of ridge regression is given by:

$$\tilde{E}(w) = \frac{1}{2}\sum_{n=1}^{N}(t_n - w^T\phi(x_n))^2 + \frac{\lambda}{2}w^Tw$$

   Show that this error function is convex and is minimized by:

$$w^* = (\lambda I + \phi^T\phi)^{-1}\phi^T t$$

   Also show that $(\lambda I + \phi^T\phi)$ is invertible for any $\lambda > 0$.

(Note 1: To simplify and keep your solution concise, use vector/matrix format (e.g., gradient, Hessian, etc.,) for your expressions.
Note 2: Here, the target vector $t \in \mathbb{R}^N$ and the matrix $\phi \in \mathbb{R}^{N\times d'}$ represents all the N input datapoints after transformation by the feature-mapping function $\phi(.) : \mathbb{R}^d \to \mathbb{R}^{d'}$. For example, the $\phi(.)$ for performing k-degree polynomial regression on a d-dimensional input for $k = 2, d = 2$ is given by $\phi([x_1, x_2]) = [1, x_1, x_2, x_1^2, x_2^2, x_1x_2]$.)

Solution:

$$\tilde{E}(w) = \frac{1}{2}\sum_{n=1}^{N}(t_n - w^T\phi(x_n))^2 + \frac{\lambda}{2}w^Tw$$

First derivative w.r.t w is:

$$\frac{\partial E}{\partial w} = 2\|t_n - w^T\phi(x_n)\|_2\left(\frac{t_n - w^T\phi(x_n)}{\|t_n - w^T\phi(x_n)\|_2}\right)(-\phi(x_n)) + \lambda\|w\|_2\left(\frac{w}{\|w\|_2}\right)$$
$$\frac{\partial E}{\partial w} = 2\left(t_n - w^T\phi(x_n)\right)(-\phi(x_n)) + \lambda w$$
$$\frac{\partial E}{\partial w} = -2\phi(x_n)^Tt_n + 2\phi(x_n)^T\phi(x_n)w + \lambda I w$$

Now for $\lambda > 0$

$$\frac{\partial^2 E}{\partial w^2} = \lambda I + 2\phi(x_n)^T\phi(x_n)$$

Thus we can conclude λI is a positive definite for any $\lambda > 0$ and $\phi(x_n)^T\phi(x_n)$ is a positive semidefinite and their total sum is positive definite then it is strictly convex.
Now for w^*, $\frac{\partial E}{\partial w} = 0$ Then, $-2\phi(x_n)^Tt_n + 2\phi(x_n)^T\phi(x_n)w + \lambda I w = 0$

$$-2\phi(x_n)^Tt_n + 2\phi(x_n)^T\phi(x_n)w + \lambda I w = 0$$
$$2\phi(x_n)^T\phi(x_n)w + \lambda I w = 2\phi(x_n)^Tt_n$$
$$\left(\phi(x_n)^T\phi(x_n) + \lambda I\right)w = 2\phi(x_n)^Tt_n$$
$$w^* = 2\left(\phi(x_n)^T\phi(x_n) + \lambda I\right)^{-1}\phi(x_n)^Tt_n$$

Since the total resultant is positive definite that is its invertible $\left(\phi(x_n)^T\phi(x_n) + \lambda I\right)$ is invertible.