Assignment 2: Linear Regression, Bayesian Inference, and Optimization

Due November 4, 2022 at 23:59

This assignment is to be done individually.

Important Note: The university policy on academic dishonesty (cheating) will be taken very seriously in this course. You may not provide or use any solution, in whole or in part, to or by another student.

You are encouraged to discuss the concepts involved in the questions with other students. If you are in doubt as to what constitutes acceptable discussion, please ask! Further, please take advantage of office hours offered by the instructor and the TA if you are having difficulties with this assignment.

DO NOT:

- Give/receive code or proofs to/from other students
- Use Google to find solutions for assignment

DO:

- Meet with other students to discuss assignment (it is best not to take any notes during such meetings, and to re-work assignment on your own)
- Use online resources (e.g. Wikipedia) to understand the concepts needed to solve the assignment.

Submitting Your Assignment

- You must submit a report in PDF format to **Gradescope**. You may typeset your assignment in LaTeX or Word, or submit neatly handwritten and scanned solutions. We will not be able to give credit to solutions that are not legible.
- Please indicate on Gradescope the area of your submission that corresponds to each question part.
- In addition to the PDF report, you must submit to **Canvas** a zip file containing three Jupyter notebooks, one for each of the questions.

---

1 Linear Regression and Model Selection

This Jupyter notebook builds on the linear regression code that we covered in the class which trains an ordinary least square on a synthesized data. For this question, you are to implement cross-validation and L2 regularization and visualize the results.

- Include all plots and descriptions in your written pdf report
- Hint: You can use matplotlib.fill_between function to visualize the variance.

1.1 Cross-Validation

Cross-validation methods involves

a) Withholding a random subset of the data during model fitting.

b) Quantifying how accurate the withheld data are predicted.

c) Repeating this process to measure prediction accuracy.

For this section, add your implementation in the linked notebook and plot the means and variances of training, validation, and test losses for polynomial degrees up to 20. Describe how results change with respect to the polynomial degree, and explain why in 2-3 sentences.

1.2 Regularization

To alleviate the issues of poor generalization, we could also use regularization. Regularization in general can be used to improve model generalization by penalizing large components of the model parameters in the loss function during training, and for large data sets cross validation may not be necessary.

For this question, we will still use cross validation to obtain more accurate validation losses for models obtained from L2 regularization. Add your implementation in the notebook and plot training, validation, and test losses with variance for different polynomial degrees for a few different penalty factors. Compare the regularization with the best hyper parameter you found to the model without regularization and explain how regularization improves the model performance.

2 MAP Estimate and Bayesian Inference

Given a simple linear model $y = wx + \epsilon$, where $x, y, w \in \mathbb{R}$ and $\epsilon \sim \mathcal{N}(0, 1)$. Suppose we know from prior knowledge that $w \sim \mathcal{N}(0, 1)$. Note that this implies $\lambda = \sigma = 1$. The training data consists of 2 data points, $(1, \frac{1}{3})$, $(3, \frac{1}{6})$.

a) Calculate w_{MAP}.

b) Calculate $p(w \mid \mathcal{D})$, where $\mathcal{D} = \{(1, \frac{1}{3}), (3, \frac{1}{6})\}$. What is the mean $\mathbb{E}[w|\mathcal{D}]$ and standard deviation σ_w of the posterior distribution $p(w \mid \mathcal{D})$?

c) Plot the two data points, the mean regression line $y = \mathbb{E}[w|\mathcal{D}]x$, and the two lines that are one-standard-deviation above and below the mean regression line, i.e. $y = (\mathbb{E}[w|\mathcal{D}] \pm \sigma_w)x$.

d) From the plot, which data point is closer to the mean regression line? Explain why in 1-2 sentences.

3 Nonlinear Optimization

In this question, we implement iterative algorithms to solve a nonlinear optimization problem. To this end, a Jupyter notebook is provided, and you should complete each designated code section according to its instructions.

Consider the following objective function.

$$f(x, y) = 4x^2 + \frac{(y+3)^2}{15} \tag{1}$$

a) Plot the objective function for the specified range of the inputs in the notebook.

b) Implement a basic Gradient Descent algorithm for the objective function (1).

- Use step size of 0.2, total iterations of 200, and initial point of $(-10, 10)$ for the objective function. Provide a contour plot of the solutions.
- Change the step size to 0.3. Report your observation.

c) Implement the Adaptive Gradient (AdaGrad) algorithm for the objective function (1).

- Use step size of 0.2, total iterations of 200, and initial point of $(-10, 10)$ for the objective function. Provide a contour plot of the solutions.
- Change the step size to 0.3. Report and plot your observation.

d) Implement the Adam algorithm for the objective function (1).

- Use step size of 0.2, total iterations of 200, and initial point of $(-10, 10)$ for the objective function. Use default values for other parameters. Provide a contour plot of the solutions.
- Does the solution converge to the minimal objective value?