

DDA3020: Assignment I

February 26, 2023

This assignment accounts for 14/100 of the final score. Homework due: **11:59 pm, March 12, 2023**

1 Written Problems (50 points)

1.1 Given the following denominator layout derivatives, (13 points)

- **Differentiation of a scalar function w.r.t. a vector:** If $f(\mathbf{w})$ is a scalar function of d variables, \mathbf{w} is a $d \times 1$ vector, then differentiation of $f(\mathbf{w})$ w.r.t. \mathbf{w} results in a $d \times 1$ vector

$$\frac{df(\mathbf{w})}{d\mathbf{w}} = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \vdots \\ \frac{\partial f}{\partial w_d} \end{bmatrix}$$

- **Differentiation of a vector function w.r.t. a vector:** If $\mathbf{f}(\mathbf{w})$ is a vector function of size $h \times 1$ and \mathbf{w} is a $d \times 1$ vector, then differentiation of $\mathbf{f}(\mathbf{w})$ w.r.t. \mathbf{w} results in a $d \times h$ matrix

$$\frac{d\mathbf{f}(\mathbf{w})}{d\mathbf{w}} = \begin{bmatrix} \frac{\partial f_1}{\partial w_1} & \cdots & \frac{\partial f_h}{\partial w_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial w_d} & \cdots & \frac{\partial f_h}{\partial w_d} \end{bmatrix}$$

Please prove the following derivatives:

Consider $\mathbf{X} \in \mathbb{R}^{h \times d}$ and $\mathbf{y} \in \mathbb{R}^{h \times 1}$, which are not functions of \mathbf{w} :

$$\frac{d(\mathbf{y}^\top \mathbf{X} \mathbf{w})}{d\mathbf{w}} = \mathbf{X}^\top \mathbf{y}, \quad (4 \text{ points})$$

$$\frac{d(\mathbf{w}^\top \mathbf{w})}{d\mathbf{w}} = 2\mathbf{w}, \quad (4 \text{ points})$$

Consider $\mathbf{X} \in \mathbb{R}^{d \times d}$ and $\mathbf{w} \in \mathbb{R}^{d \times 1}$ (5 points):

$$\frac{d(\mathbf{w}^\top \mathbf{X} \mathbf{w})}{d\mathbf{w}} = (\mathbf{X} + \mathbf{X}^\top) \mathbf{w}$$

1.2 Suppose we have training data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}, i = 1, 2, \dots, N$. Consider $f_{\mathbf{w}, b}(\mathbf{x}) = \mathbf{x}^\top \mathbf{w} + b$. (12 points)

(1) Find the closed-form solution of the following problem

$$\min_{\mathbf{w}, b} \sum_{i=1}^N (f_{\mathbf{w}, b}(\mathbf{x}_i) - y_i)^2 + \lambda \bar{\mathbf{w}}^\top \bar{\mathbf{w}}, \quad (1)$$

where $\bar{\mathbf{w}} = \hat{\mathbf{I}}_d \mathbf{w} = [0, w_1, w_2, \dots, w_d]^\top$. (6 points)

(2) Show how to use gradient descent to solve the problem. (6 points)

1.3 Prove that:

- (1) $f(x) = x^2$ is convex. (4 points)
- (2) every affine function $f(x) = ax + b$ is convex, but not strictly convex; (4 points)
- (3) $f(x) = |x|$ is convex, but not strictly convex; (5 points)

1.4 Suppose x_1, x_2, \dots, x_N are drawn from $Laplace(\mu, b)$. Calculate the MLE (maximum likelihood estimation) of μ and b . Hint: Use the logarithmic trick to process multiple exponential items. (12 points)

2 Programming (50 points)

The boston.csv file contains Boston Housing Dataset. The Boston Housing Dataset is a derived from information collected by the U.S. Census Service concerning housing in the area of Boston MA. The following describes the dataset columns:

- CRIM - per capita crime rate by town
- ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
- INDUS - proportion of non-retail business acres per town.
- CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
- NOX - nitric oxides concentration (parts per 10 million)
- RM - average number of rooms per dwelling
- AGE - proportion of owner-occupied units built prior to 1940
- DIS - weighted distances to five Boston employment centres
- RAD - index of accessibility to radial highways
- TAX - full-value property-tax rate per \$10,000
- PTRATIO - pupil-teacher ratio by town
- B - $1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town
- LSTAT - % lower status of the population
- MEDV - Median value of owner-occupied homes in \$1000's

You need to use appropriate attributes in 'crim, zn, indus, chas, nox, rm, age, dis, rad, tax, ptratio, b, lstat' to predict the last attributes 'MEDV'. You need to finish the following steps:

- **Step 1:** use **pandas** library to check the data in the dataset. Process incomplete data point such as 'NaN' or 'Null'. Briefly summarize the characteristics of this dataset and guess which is the most relevant attribute for MEDV.
- **Step 2:** use **seaborn** library to visualize the dataset. Plot the MEDV distributions over each attribute. Briefly analyze the characteristics of the attributes and revise the assumption in Step 1 if necessary.
- **Step 3:** use **seaborn.heatmap** function to plot the pairwise correlation on data. Select the good attributes which are good indications of using as predictors. Report your findings.
- **Step 4:** use **sklearn.preprocessing.MinMaxScaler** function to scale the columns you select in Step 3. Then use **seaborn.regplot** to plot the relevance of these columns against MEDV with 95% confidence interval.

- **Step 5:** Randomly split the data into two parts, one contains 80% of the samples and the other contains 20% of the samples. Use the first part as training data and train a linear regression model and make prediction on the second part with gradient descent methods. X should be the attributes you select in previous steps.

Report the training error and testing error in terms of RMSE. Plot the loss curves in the training process. **Notice: you need to write the codes of learning the parameters by yourself. Do not use the regression packages of Sklearn.**

- **Step 6:** Repeat the splitting, training, and testing for 10 times with different parameters such as step size, iteration steps, etc. Use a loop and print the RMSEs in each trial. Analyze the influence of different parameters on RMSE.

2.1 Submission Format

- code.ipynb (without Input Data included) - 25 pts - your jupyter notebook files should contain the running output of each step (numbers, plots, etc.). If your notebook has only code but no output results, you will get a discounted score.
- Submit report containing linear model description, loss function, hyperparameter settings, RMSE equation, outputs (errors, plots, figures), and relevant analysis required in above steps. This should be included as part of written assignment in pdf file. - 25 pts.
- Note: Please include all results from your model in the report. You will receive no credits if we have to run the code to get outputs. The recommended length of the report is about 3-5 pages. **If the report is too short, the score will be deducted for lacking sufficient contents. There is also no credit bonus for too long reports.** The overall submission format is one **ipynb** file and one **pdf** file containing both answers for written and report.