

DDA3020: Assignment I

September 29, 2022

Homework due: **11:59 pm, Oct 17, 2022**. The weight of this assignment among all assignments is **13/60**.

1 Written Problems (50 points)

1.1 Given the following denominator layout derivatives, (15 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 $d \times h$ vector

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

Please prove the following derivatives, and \mathbf{X} and \mathbf{y} are not functions of \mathbf{w} :

$$\begin{aligned} \frac{d(\mathbf{X}^\top \mathbf{w})}{d\mathbf{w}} &= \mathbf{X}, \\ \frac{d(\mathbf{y}^\top \mathbf{X} \mathbf{w})}{d\mathbf{w}} &= \mathbf{X}^\top \mathbf{y} \\ \frac{d(\mathbf{w}^\top \mathbf{X} \mathbf{w})}{d\mathbf{w}} &= (\mathbf{X} + \mathbf{X}^\top) \mathbf{w} \end{aligned}$$

1.2 Suppose we have training data $\{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{y}_i \in \mathbb{R}^k$, $i = 1, 2, \dots, N$. (1) Find the closed-form solution of the following problem

$$\min_{\mathbf{W}, \mathbf{b}} \sum_{i=1}^N \alpha_i \|\mathbf{y}_i - \mathbf{W} \mathbf{x}_i - \mathbf{b}\|^2,$$

where $\alpha_1, \dots, \alpha_n$ are weights for different samples; (2) Show how to use gradient descent to solve the problem. (12 points)

1.3 Prove that: (1) $f(x) = x^4$ is convex; (2) $f(x) = |x|$ is convex; (3) $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ is convex, where \mathbf{A} is a matrix. (12 points)

1.4 Suppose x_1, x_2, \dots, x_N are drawn from $\mathcal{N}(\mu, \sigma^2)$. Show that the MLE (maximum likelihood estimation) of σ^2 is $\sigma_{ML}^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{ML})^2$. (11 points)

2 Programming (50 points)

2.1. The CSV or XLS file contains a dataset for regression. There are 7750 samples with 25 features (described in the doc file). This data is for the purpose of bias correction of next-day maximum and minimum air temperatures forecast of the LDAPS model operated by the Korea Meteorological Administration over Seoul, South Korea. This data consists of summer data from 2013 to 2017. The input data is largely composed of the LDAPS model's next-day forecast data, in-situ maximum and minimum temperatures of present-day, and geographic auxiliary variables. There are two outputs (i.e. next-day maximum and minimum air temperatures) in this data. Hindcast validation was conducted for the period from 2015 to 2017.

You need to delete the first two attributes (station and date), and use attributes 3-23 to predict attributes 24 and 25. 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. Report the training error and testing error in terms of RMSE.

Repeat the splitting, training, and testing for 10 times. Use a loop and print the RMSEs in each trial.

2.2. The classification data file contains the iris dataset. This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other. There are five attributes: 1. sepal length in cm; 2. sepal width in cm; 3. petal length in cm; 4. petal width in cm; 5. class: – Iris Setosa – Iris Versicolour – Iris Virginica. You need to use the first attributes to predict the last attribute, namely, classifying the data into three classes. 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 model and make classification on the second part. Report the training error and testing error in terms of classification error rate (number of miss-classified samples divided by number of all samples)

Repeat the splitting, training, and testing for 10 times. Use a loop and print the classification errors in each trial.

Note that you need to write the codes of learning the parameters by yourself. Do not use the classification or regression packages of Sklearn. You can use their tools to shuffle the data randomly for splitting.