DDA3020 Homework 1

Due date: Oct 14, 2024

Instructions

- The **deadline** is 23:59, Oct 14, 2024.
- The weight of this assignment in the final grade is 20%.
- **Electronic submission**: Turn in solutions electronically via Blackboard. Be sure to submit your homework as one pdf file plus two python scripts. Please name your solution files as "DDA3020HW1_studentID_name.pdf", "HW1_yourID_Q1.ipynb" and "HW1_yourID_Q2.ipynb". (.py files also acceptable)
- Note that **late submissions** will result in discounted scores: 0-24 hours → 80%, 24-120 hours → 50%, 120 or more hours → 0%.
- Answer the questions in English. Otherwise, you'll lose half of the points.
- Collaboration policy: You need to solve all questions independently and collaboration between students is **NOT** allowed.

1 Written Problems (50 points)

1.1. (Learning of Linear Regression, 25 points) Suppose we have training data:

$$\{(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2), \ldots, (\boldsymbol{x}_N, \boldsymbol{y}_N)\},$$

where $\boldsymbol{x}_i \in \mathbb{R}^d$ and $\boldsymbol{y}_i \in \mathbb{R}^k$, $i = 1, 2, \ldots, N$.

i) (9 pts) Find the closed-form solution of the following problem.

$$\min_{\boldsymbol{W}, \boldsymbol{b}} \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{b}\|_2^2,$$

ii) (8 pts) Show how to use gradient descent to solve the problem. (Please state at least one possible Stopping Criterion)

iii) (8 pts) We further suppose that $x_1, x_2, \ldots, x_N$ are drawn from $\mathcal{N}(\mu, \sigma^2)$. Show that the maximum likelihood estimation (MLE) of σ^2 is $\hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{MLE})^2$.

1.2. (Support Vector Machine, 25 points) Given two positive samples $x_1 = (3,3)^T$, $x_2 = (4,3)^T$, and one negative sample $x_3 = (1,1)^T$, find the maximum-margin separating hyperplane and support vectors.
Solution steps:

i) Formulating the Optimization Problem (5 pts)

ii) Constructing the Lagrangian (5 pts)

iii) Using KKT Conditions (5 pts)

iv) Solving the Equations (5 pts)

v) Determining the Hyperplane Equation and Support Vectors (5 pts)

2 Programming (50 points)

2.1. (Linear regression, 25 points) We have a labeled dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_n, y_n)\}$, with $\mathbf{x}_i \in \mathbb{R}^d$ being the d-dimensional feature vector of the i-th sample, and $y_i \in \mathbb{R}$ being real valued target (label).

A linear regression model is give by

$$f_{w_0,\ldots,w_d}(\mathbf{x}) = w_0 + w_1x_1 + w_2x_2 + \cdots + w_dx_d, \tag{1}$$

where w_0 is often called bias and $w_1, w_2, \ldots, w_d$ are often called coefficients.

Now, we want to utilize the dataset $\mathcal{D}$ to build a linear model based on linear regression. We provide a training set $\mathcal{D}_{\text{train}}$ that includes 2024 labeled samples with 11 features (See linear_regression_train.txt) to fit model, and a test set $\mathcal{D}_{\text{test}}$ that includes 10 unlabeled samples with 11 features (see linear_regression_test.txt) to estimate model.

1. Using the LinearRegression class from Sklearn package to get the bias w_0 and the coefficients $w_1, w_2, \ldots, w_{11}$, then computing the $\hat{y} = f(\mathbf{x})$ of test set $\mathcal{D}_{\text{test}}$ by the model trained well. (Put the estimation of $w_0, w_1, \ldots, w_{11}$ and these $\hat{y}$ in your answers.)

2. Implementing the linear regression by yourself to obtain the bias w_0 and the coefficients $w_1, w_2, \ldots, w_{11}$, then computing the $\hat{y} = f(\mathbf{x})$ of test set $\mathcal{D}_{\text{test}}$. (Put the estimation of $w_0, w_1, \ldots, w_{11}$ and these $\hat{y}$ in your answers. It is allowed to compute the inverse of a matrix using the existing python package.)

1.1. (Learning of Linear Regression, 25 points) Suppose we have training data:

$$\{(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2), \ldots, (\boldsymbol{x}_N, \boldsymbol{y}_N)\},$$

where $\boldsymbol{x}_i \in \mathbb{R}^d$ and $\boldsymbol{y}_i \in \mathbb{R}^k$, $i = 1, 2, \ldots, N$.

i) (9 pts) Find the closed-form solution of the following problem.

$$\min_{\boldsymbol{W}, \boldsymbol{b}} \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{b}\|_2^2,$$

ii) (8 pts) Show how to use gradient descent to solve the problem. (Please state at least one possible Stopping Criterion)

i)

$$J(W) = \sum_{i=1}^{N} \| y_i - W x_i - b \|_2^2$$

$$= (y - Wx - b)^T (y - Wx - b)$$

$$= (y^T - x^T W^T - b^T)(y - Wx - b)$$

$$= (\cancel{y^T y} - y^T W x - \cancel{y^T b} - y x^T W^T +$$

$$x^T W^T W x + x^T W^T b - \cancel{b^T y}$$

$$+ b^T W x + \cancel{b^T b})$$

· Set $\frac{\partial}{\partial W} J(W) = 0$:

$$0 = -y^T x - y x^T + 2 x^T x W + x^T b + b^T x$$

$$2(y^T x - x^T b) = 2 x^T x W$$

$$W = (x^T x)^{-1} (y^T x - x^T b)$$

ii) Gradient descent $W \leftarrow W - \alpha \cdot \frac{\partial J(W, b)}{\partial W}$, $b \leftarrow b - \lambda \cdot \frac{\partial J(W, b)}{\partial b}$, where α, λ are learning rates

let cost function $J(W, b) = \frac{1}{2N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$, where $\hat{y}_i = W_i x_i + b$

$$\Rightarrow J(W, b) = \frac{1}{2N} \sum_{i=1}^{N} (y_i - (W x_i + b))^2$$

$$\frac{\partial J}{\partial W} = \frac{1}{N} \sum_{i=1}^{N} (-x_i)(y_i - (W x_i + b))$$

At min point: $\frac{\partial J}{\partial W} = 0$, $0 = \sum_{i=1}^{N} (-x_i y_i + W x_i^2 + x_i b)$

Rewriting: $\sum_{i=1}^{N} w x_i + \sum_{i=1}^{N} x_i b = \sum_{i=1}^{N} x_i y_i$ ——— (1)

Now partial derivative w.r.t. 'b'

$$\frac{\partial J}{\partial b} = \frac{1}{N} \sum_{i=1}^{N} -\left(y_i - (w x_i + b)\right)$$

At min point: $\frac{\partial J}{\partial b} = 0$

$$0 = \sum_{i=1}^{N} (-y_i + w x_i + b) \iff \sum_{i=1}^{N} y_i = w \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} b$$

$$\iff \sum_{i=1}^{N} y_i = w \sum_{i=1}^{N} x_i + Nb \quad \text{———} \quad (2)$$

With the partial derivatives $\frac{\partial J}{\partial w}$ and $\frac{\partial J}{\partial b}$, we can update our weight & bias using the formulas:

$$\left.\begin{array}{ll} (*) & w \leftarrow w - \alpha \frac{\partial J}{\partial w} \\ \\ (**) & b \leftarrow b - \lambda \frac{\partial J}{\partial b} \end{array}\right\} \text{This is the gradient descent looping.}$$

Stopping criterion:

1. Max iterations: # of loops exceed a defined threshold.

2. change in cost b/t iterations is less than some small threshold ε.

   i.e. $\| \nabla f(w,b) \| < \varepsilon \iff | f(w,b)_{prev} - f(w,b)_{current} | < \varepsilon$.

iii) $L = (\mu, \sigma^2 \mid x_1 \ldots x_n) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2)$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}^{\,n}}\right) e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\ell(\mu, \sigma^2) = 0 - \frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= -\frac{n}{2}\ln 2 - \frac{n}{2}\ln\pi - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

To obtain $\hat{\sigma}^2_{MLE}$, take partial derivative wrt to σ^2

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = 0 - \frac{n}{2\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2 \cdot \frac{1}{(\sigma^2)^2}$$

At max, $\frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = 0$

$$\Rightarrow \quad 0 = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$0 = n\sigma^2 - \sum_{i=1}^{n}(x_i-\mu)^2$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\therefore \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_{MLE})^2 \quad \text{(shown)}$$

1.2. (Support Vector Machine, 25 points) Given two positive samples $x_1 = (3,3)^T$, $x_2 = (4,3)^T$, and one negative sample $x_3 = (1,1)^T$, find the maximum-margin separating hyperplane and support vectors.

Solution steps:

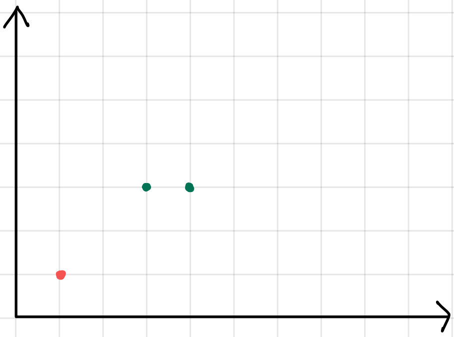
i) Formulating the Optimization Problem (5 pts)

ii) Constructing the Lagrangian (5 pts)

iii) Using KKT Conditions (5 pts)

iv) Solving the Equations (5 pts)

v) Determining the Hyperplane Equation and Support Vectors (5 pts)



i) Hyperplane: $w^T x + b = 0$

impose constraints on x_i's, $i = 1, 2, 3$

+ve: $w^T x_1 + b \geq 1$

$w^T x_2 + b \geq 1$

-ve: $w^T x_3 + b \leq -1$

Then, $y_i(w^T x_i + b) \geq 1, \ \forall i$, with margin $\frac{1}{\|w\|}$

$\therefore$ obj function: $\min_{w,b} \frac{1}{2}\|w\|^2$ s.t. $y_i(w^T x_i + b) \geq 1, \ \forall i$ $\iff$ $\min_{w,b} \frac{1}{2}\|w\|^2$ s.t. $1 - y_i(w^T x_i + b) \leq 0, \ \forall i$

ii) The Lagrangian: $\mathcal{L}(w, b, \alpha) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i \left(1 - y_i(w^T x_i + b)\right)$

substituting the datapoints: $\{(x_1 = (3,3)^T, y_1 = 1), (x_2 = (4,3)^T, y_2 = 1), (x_3 = (1,1)^T, y_3 = -1)\}$

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}\|w\|^2 + \alpha_1 - \alpha_1(w^T x_1 + b) + \alpha_2 - \alpha_2(w^T x_2 + b) + \alpha_3 + \alpha_3(w^T x_3 + b)$$

where $x_1 = (3,3)^T$, $x_2 = (4,3)^T$, $x_3 = (1,1)^T$

iii) The primal & dual optimal solutions should satisfy KKT conditions

Stationarity: $\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{2}(2w) - \alpha_1 x_1 - \alpha_2 x_2 + \alpha_3 x_3 = w - \alpha_1 x_1 - \alpha_2 x_2 + \alpha_3 x_3 = 0$

$\Rightarrow w = \alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3 = \alpha_1 (3,3)^T + \alpha_2 (4,3)^T - \alpha_3 (1,1)^T$ (*)

$\frac{\partial \mathcal{L}}{\partial b} = -\alpha_1 - \alpha_2 + \alpha_3 = 0 \iff \alpha_1 + \alpha_2 - \alpha_3 = 0$

$\iff \alpha_1 + \alpha_2 = \alpha_3$ (**)

Feasibility: $\alpha_1 \geq 0$, $1 - w^T x_1 - b \leq 0 \iff 1 - w^T (3,3)^T - b \leq 0$

$\alpha_2 \geq 0$, $1 - w^T x_2 - b \leq 0 \iff 1 - w^T (4,3)^T - b \leq 0$

$\alpha_3 \geq 0$, $1 + w^T x_3 + b \leq 0 \iff 1 + w^T (1,1)^T + b \leq 0$

Complementary slackness:

$$\alpha_1 - \alpha_1 W^T x_1 - \alpha_1 b = 0 \iff \alpha_1 - \alpha_1 W^T (3,3)^T - \alpha_1 b = 0$$

$$\alpha_2 - \alpha_2 W^T x_2 - \alpha_2 b = 0 \iff \alpha_2 - \alpha_2 W^T (4,3)^T - \alpha_2 b = 0$$

$$\alpha_3 + \alpha_3 W^T x_3 + \alpha_3 b = 0 \iff \alpha_3 - \alpha_3 W^T (1,1)^T - \alpha_3 b = 0$$

iv) solving the Eqns:

From the feasibility conditions, we have the equalities

$$\left.\begin{aligned} &1:\ W^T(3,3)^T + b = 1 \\ &2:\ W^T(4,3)^T + b = 1 \\ &3:\ W^T(1,1)^T + b = -1 \end{aligned}\right\} \text{Sub } W \text{ into these eqns}$$

For 1:
$$\left[\alpha_1 (3,3)^T + \alpha_2 (4,3)^T - \alpha_3 (1,1)^T\right]^T (3,3)^T + b = 1$$

$$\left[\begin{pmatrix} 3\alpha_1 \\ 3\alpha_1 \end{pmatrix} + \begin{pmatrix} 4\alpha_2 \\ 3\alpha_2 \end{pmatrix} - \begin{pmatrix} \alpha_3 \\ \alpha_3 \end{pmatrix}\right]^T \begin{pmatrix} 3 \\ 3 \end{pmatrix} + b = 1$$

$$\begin{bmatrix} 3\alpha_1 + 4\alpha_2 - \alpha_3 \\ 3\alpha_1 + 3\alpha_2 - \alpha_3 \end{bmatrix}^T \begin{pmatrix} 3 \\ 3 \end{pmatrix} + b = 1$$

$$9\alpha_1 + 12\alpha_2 - 3\alpha_3 + 9\alpha_1 + 9\alpha_2 - 3\alpha_3 + b = 1$$

$$18\alpha_1 + 21\alpha_2 - 6\alpha_3 + b = 1 \quad \text{———} \quad 4$$

For 2:
$$\begin{bmatrix} 3\alpha_1 + 4\alpha_2 - \alpha_3 \\ 3\alpha_1 + 3\alpha_2 - \alpha_3 \end{bmatrix}^T \begin{pmatrix} 4 \\ 3 \end{pmatrix} + b = 1$$

$$12\alpha_1 + 16\alpha_2 - 4\alpha_3 + 9\alpha_1 + 9\alpha_2 - 3\alpha_3 + b = 1$$

$$21\alpha_1 + 25\alpha_2 - 7\alpha_3 + b = 1 \quad \text{———} \quad 5$$

For 3.
$$\begin{bmatrix} 3\alpha_1 + 4\alpha_2 - \alpha_3 \\ 3\alpha_1 + 3\alpha_2 - \alpha_3 \end{bmatrix}^T \begin{pmatrix} 1 \\ 1 \end{pmatrix} + b = -1$$

$$6\alpha_1 + 7\alpha_2 - 2\alpha_3 + b = -1 \quad \text{———} \quad 6$$

Taking ④ $- 3\times$ ⑥ : $\quad b - 3b = 1 - (-3)$

$$-2b = 4$$
$$b = -2$$

Taking ⑤ $- 3\times$ ⑥ : $\quad 3\alpha_1 + 4\alpha_2 - \alpha_3 - 2b = 1 - (-3)$

Sub in $\alpha_3 = \alpha_1 + \alpha_2$, $\quad b = -2$

$$3\alpha_1 + 4\alpha_2 - \alpha_1 - \alpha_2 + 4 = 4$$

$$2\alpha_1 + 3\alpha_2 = 0 \quad \text{———} \quad ⑦$$

Sub in $\alpha_3 = \alpha_1 + \alpha_2$, $b = -2$ into ⑤:

$$21\alpha_1 + 25\alpha_2 - 7\alpha_1 - 7\alpha_2 - 2 = -1$$

$$14\alpha_1 + 18\alpha_2 = 3 \quad \text{———} \quad ⑧$$

Solving ⑦ and ⑧ : $\quad \alpha_1 = 3/2, \; \alpha_2 = -1$

$$\therefore \; \alpha_3 = 1/2$$

and $\quad w = \frac{3}{2}\begin{pmatrix} 3 \\ 3 \end{pmatrix} - \begin{pmatrix} 4 \\ 3 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$= \begin{pmatrix} 9/2 \\ 9/2 \end{pmatrix} - \begin{pmatrix} 4 \\ 3 \end{pmatrix} - \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

$$= \begin{pmatrix} 9/2 - 9/2 \\ 9/2 - 7/2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

v)

Hyperplane Eqn. $\quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} x_i - 2 = 0$

The support vectors : $\quad x_1 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$