DDA3020 Homework 1

Due date: March 09, 2025

Instructions

- The **deadline** is 23:59, March 09, 2025.
- The weight of this assignment in the final grade is 20%.
- **Electronic submission**: Turn in solutions electronically via Blackboard. Be sure to submit your answers as one pdf file plus two python scripts for programming questions. Please name your solution files as "DDA3020HW1_studentID_name.pdf", "HW1_name_Q1.ipynb" and "HW1_name_Q2.ipynb" (".py" files are also acceptable).
- The complete and executable codes must be submitted. If you only fill in some of the results in your answer report for programming questions and do not submit the source code (.py or .ipynb files), you will receive 0 points for the question.
- Note that **late submissions** will result in discounted scores: 0-48 hours $\rightarrow$ 50%, more hours $\rightarrow$ 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. (Closed-Form Solution, 25 points) Given a ridge regression with data $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{y} \in \mathbb{R}^n$, where n is the number of data, m is the number of attributes used for prediction, under the assumption that data $\mathbf{X}$ is centered, i.e., $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$, show that the closed-form solution of the cost function

$$J(\mathbf{w}, w_0) = (\mathbf{y} - \mathbf{X}\mathbf{w} - w_0\mathbf{1})^T(\mathbf{y} - \mathbf{X}\mathbf{w} - w_0\mathbf{1}) + \lambda \mathbf{w}^T\mathbf{w}$$

is

$$\hat{w}_0 = \bar{\mathbf{y}}$$

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$$

1.2. (Support Vector Machine, 25 points) Consider a dataset with 2 points in 1D: ($x_1 = 0, y_1 = -1$) and ($x_2 = 1, y_2 = 1$). Consider mapping each point to 3D using the feature vector $\boldsymbol{\phi}(x) = [1, \sqrt{2}x, x^2]^\top$. (This is equivalent to using a second-order polynomial kernel.) The max margin classifier has the form:

$$\min \|\boldsymbol{w}\|^2 \quad \text{s.t.}$$

$$y_1 \left(\boldsymbol{w}^T \boldsymbol{\phi}(x_1) + w_0\right) \geq 1$$

$$y_2 \left(\boldsymbol{w}^T \boldsymbol{\phi}(x_2) + w_0\right) \geq 1$$

(i) Write down a vector that is parallel to the optimal vector $\boldsymbol{w}$.

(ii) What is the value of the margin achieved by this $\boldsymbol{w}$? Hint: recall that the margin is the distance from each support vector to the decision boundary.

(iii) Solve for $\boldsymbol{w}$, using the fact that the margin is equal to $1/\|\boldsymbol{w}\|$.

(iv) Solve for w_0 using your value for $\boldsymbol{w}$ and the optimization problem above. Hint: the points will be on the decision boundary, so the inequalities will be tight, think about the geometry of 2 points in space, with a line separating one from the other.

(v) Write down the discriminant function $f(x) = w_0 + \boldsymbol{w}^\top \boldsymbol{\phi}(x)$ as an explicit function of x.

2 Programming (50 points)

2.1. (Logistic Regression, 25 points) In this question, we will implement multi-class logistic regression using gradient descent. Follow code template provided in *logistic_assignment1.ipynb* to solve this question. Noticably, this question should not be solved by using *sklearn* package, aiming to promote your understanding on gradient descent. A mathematical derivation of gradient is required in this problem. Provide your solution in your written answers.

2.2. (Support Vector Machine, 25 points) In this question, we will explore the use of Support Vector Machines (SVM) for both linear and non-linear tasks using the *sklearn* library, as outlined in the *SVM_assignment1.ipynb* file. By following the instructions in the notebook, we will implement both types of SVMs to gain a foundational understanding of how to apply SVMs.