COMP3020 : Machine Learning

Assignment 02

2024-10-10

Instructions

- Submit your assignment via Canvas.
- Your submission should include the zip file only (code + written part)
- You may use LaTex or Word to create the written part of your assignment, but you must submit it in PDF format.
- You should be able to finish this assignment within 10 hours

1 Perceptron (20 points)

In this problem, we will work on a dataset where input and output are binary. The output of the Perceptron model is $\text{sign}(w^\intercal x + b)$, where $w \in \mathbb{R}^2, b \in \mathbb{R}$. The initial weights and bias are $w_0 = [0, -1], b_0 = \frac{1}{2}$.

Exercise 1a (10) For a sample x, it contains two feature $x_1, x_2 \in \{0, 1\}$, and has two possible outputs. Assuming we have 4 samples as follows

x_1	x_2	y
0	0	-1
0	1	1
1	0	1
1	1	1

Table 1: Features x_i and label y

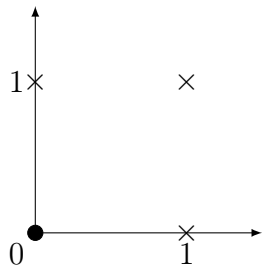

Figure 1: Illustration of the dataset, ●, × denotes samples with label $-1, 1$, respectively.

Can we learn a model that can give correct outputs for all 4 samples? If possible, show how to obtain w and b step-by-step, otherwise prove that it is impossible.

Exercise 1b (10). Now consider a different dataset as in Table 2. Can we learn a model that can give correct outputs for all 4 samples? If possible, show how to obtain w and b step-by-step, otherwise prove that it is impossible.

x_1	x_2	y
0	0	-1
0	1	1
1	0	1
1	1	-1

Table 2: Features x_i and label y

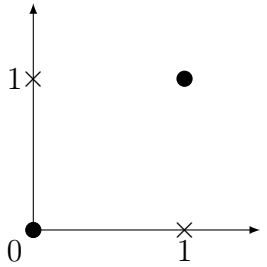

Figure 2: Illustration of the dataset, ●, × denotes samples with label $-1, 1$, respectively.

2 Linear Regression (40 points)

We are now going to learn a simple linear regression model. Given a data matrix $X \in \mathbb{R}^{n\times m}$ be the data matrix where the i-th row x_i is the feature of the $i-$th samples, an output vector $y \in \mathbb{R}^n$ where y_i is the output of x_i, we want to find a weight $w \in \mathbb{R}^m$ such that $w^\intercal x_i$ well approximates y_i.

We define an objective where the loss function is the sum of square errors between the output of our model $w^\intercal x_i$ and the true output y_i

$$L(w) = \sum_{i=1}^{n} |w^\intercal x_i - y_i|^2 = \|Xw - y\|^2. \tag{1}$$

Exercise 2a (10). Compute the derivative of $L(w)$ with respect to w.

Exercise 2b (10). Assume that X is full column rank, find the weight w^* where the derivative of $L(w^*)$ is zero.

Exercise 2c (10). Assuming an inductive bias of simple solutions, we add an L_2 regularization to the objective function

$$L_{\text{ridge}} = \|Xw - y\|^2 + \lambda\|w\|_2^2, \quad \text{where } \lambda > 0$$

Compute the derivative of L_{ridge} with respect to w. What value of λ has a unique w^* that makes the derivative equal to 0, and what is w^* in that case?

Exercise 2d (10). Show that we can obtain L_2 regularization by adding artificial samples to the dataset and training with ordinary least square regression. Describe your method for creating these artificial samples, particularly their features and labels.

3 Coding Questions (40 points)

In this section, you will implement Perceptron and Linear Regression algorithms with the formula that we have derived so far. The goal of these exercises is to help you get used to the training process of simple models and evaluate them. For the first two questions, you need to train these models using different algorithms and test the performance. For the last question, you will train a model for a text semantic classification task and submit the output to Kaggle `https://www.kaggle.com/t/24eae58e93964645985d075969ab1eb8`.

You will submit the solution for written questions and coding questions in a single *zip* file. The deadline on Kaggle is one hour after the deadline on Canvas. After that, your Kaggle submission cannot be seen and graded.

Importante note: You are allowed to create only 5 submissions per day on Kaggle, so use your time wisely!

Implementing the Perceptron algorithm	10 points
Implementing the Linear Regression algorithm	15 points
Now, go all out and train a Classification Model on text dataset and submit the output to Kaggle (Note that, your algorithms are restricted to kNN, Decision Trees, Perceptron).	15 points
Bonus point for 5 students with the highest scores on Kaggle leaderboard	5 points

Table 3: Classification Grading Rubrics