

CISC 484
Homework 1
Due 11:59pm Thursday March 9, 2023 on Canvas

Part 1: This part must be done individually.

Q1: Assume that a training set has been provided where the target value is specified by the target function $f(x) = 7x_1 + 3x_2 - 5x_3 - 1$. We wish to consider applying incremental linear regression training with initial values $w_0 = 0.5$ and $w_1 = 2$, $w_2 = 4$, $w_3 = 1$ and $\eta = 0.1$.

Assume the first training instance is given by $\langle 1, 2, 3 \rangle$. If you apply the training algorithm on this one data point, what will the new weights be?

Q2. (Logistic Regression)

I surveyed the students who took CISC 484 last year and asked them about the number of hours they spent preparing for the final exam. Based on their answers and their grades in the exam, I used Logistic Regression to predict the probability of getting an A given the hours they spent in preparation. The resulting model had $w_0 = 0.32$ and $w_1 = 1.2$.

- a. How does the odds of getting an A after studying for 6 hours compare with the odds of getting an A with 5 hours of study? (Hint. Consider the ratio between the two odds.)
- b. What is the probability of getting an A after studying for 5 hours?

Part 2: This part must be completed with your group members.

Q3. Linear Regression

Consider the following data for inducing a linear regressor.

x1	X2	X3	Target value
2	1	3	6
5	2	7	11
2	-2	4	0
-3	1	-2	0
-5	3	-1	1

Implement a Batch (= training set) linear regression training algorithm. Train it on the above dataset. Start with $w_0 = 0.1$, $w_1 = 0.2$, $w_2 = 0.1$ and $w_3 = 0.1$. Assume training rate η of 0.1. Train for

10 epochs or training error of 0, whichever comes first. Report the training error after each epoch.

Please provide the code file.

Q4. Implement linear regression algorithm (batch=training set) for the dataset containing 8 points. Four data points have output value of 0 and are given by w_1 values of -1,-2,-3 and -4. Similarly there are four data points have output value of 1 and are given by w_1 values of 1,2,3 and 4. Assume initial weights of $w_0=0.1$ and $w_1=0.1$. Assume learning rate η of 0.2.

Train for 20 epochs or if error rate difference between two consecutive iterations (epochs) is less than 1.

Draw a graph indicating the eight training points and what the induced model looks like. On the same graph, show any linear separator that separates the instances with output value of 0 from the instances with output value of 1.

Also include the code file.

Q5. Logistic Regression.

This involves using the Pima Indian dataset (all the data can be found in files with extension .pkl) and the logistic regression program on scikit-learn (see the readme file). See the HW1_Q5.zip file for the data files for this question.

1. Train on train_1 file and test on test_1 file. Which attribute/feature seems to have the greatest impact on the prediction?

Provide the weights and Wald values as well as output predictions for each test instance.

2. Now train of train_2 file and test on test_2 file. This time, we have added a new feature and retained the others with the same values. Observe the accuracies of the model remain the same as in part a. Examine the data and the model parameters. What do you observe about the first (the new feature) feature's weight and that of others. Explain what you observe and why you obtain the same predicted values.

Provide the weights and Wald values as well as output predictions for each test instance.

3. Repeat a similar exercise as (b) on train_3 and test_3. These two sets have a new feature added to data from train_1 and test_1. (We are no longer concerned about the new feature in (b) and train_2 and test_2 datasets).