

CS 3346a — CS 3121a
Assignment 4

1. Write a program to learn the linear regression model for the data supplied below. Do not use a library to calculate the gradient descent. The program is to be written in Python. Plot the data and show the linear regression line that is found. Describe how well the linear regression model models the data. Be precise in your description (for instance, using a sum of squared errors). Also, if the model doesn't model the data well, qualitatively describe a function that might do it better.

The data is in the following file: Assignment 4 - Question 1 data.xlsx (x and y data)

2. Write a program to implement logistic regression to learn to classify the training data supplied below. Do not use a library to calculate the gradient descent. The program is to be written in Python. Test the logistic regression model on the test set. Do the results seem reasonable? Be precise in your answer. You will want to generate a confusion matrix and compute the precision, recall, and F1-score.

The training data can be found here: Assignment 4 - Question 2 breast-cancer-wisconsin.training.data.txt

The test data can be found here: Assignment 4 - Question 2 breast-cancer-wisconsin.test.data.txt

Each example has 10 features, the first feature can be ignored (it is an ID number) leaving 9 features.

The last number in each example is the example label: 2=benign, 4=malignant

3. Write a program to learn the decision tree model (with attribute choice based on information gain) for the training data supplied below. Do not use a library to create the decision tree model. The program is to be written in Python. Test the decision tree model on the test set. Do the results seem reasonable? Be precise in your answer. You will want to generate a confusion matrix and compute the precision, recall, and F1-score.

The training data is located here: Assignment 4 - Question 3 training data.csv

The test data is located here: Assignment 4 - Question 3 test data.csv

There are 8 features and 1 label (hired, i.e., the final value) for each example.

For each question, hand in a file containing the Python program and a file containing the required answers. Use descriptive file names.