

Homework Assignment 3

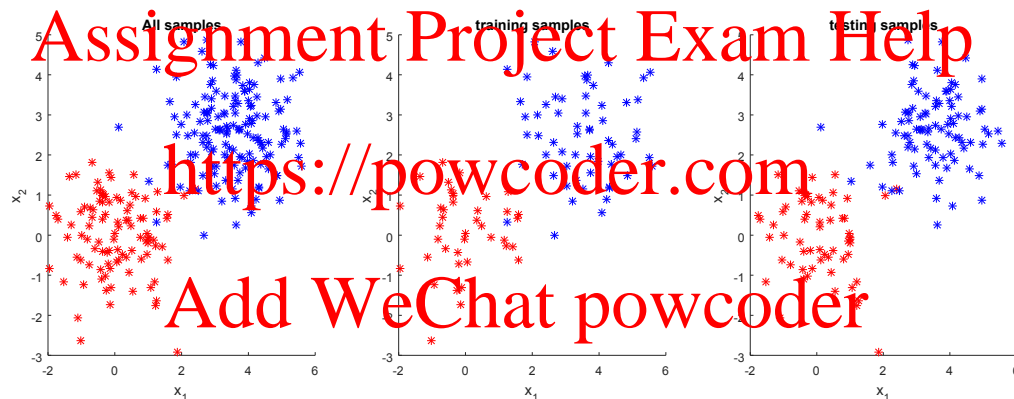
This assignment includes three problems for practicing logistic regression methods and evaluation metrics.

Problem I. Logistic Regression

Overview. The goal of this mini-project is to predict the binary class label for any sample described with two features. You will be instructed to use Logistic Regression (LR) method for solving a simulated problem.

Sample Codes. The file “main_part1.py” provides sample codes to start with. There are four major steps: generate dataset, training LR model, testing, and evaluating the prediction results.

The **first** step is to generate data (provided) and split it into training and testing subsets. You will need to write your codes to do the split. Then, the code will display the splitting results as the following figures:



The **second** step is to train a logistic regression model using the training data. To do so, you will need to use the functions we provided in the folder ‘codeLogit’. Remember there are two different implementations. **Please try both methods in this placefolder and report their performance differences.**

The **third** step is to apply the learned model to get the binary classes of testing samples. This step should be modified according to the implementation of the second step.

The **fourth** step is to compare the predictions with the ground-truth labels and calculate average errors and standard deviation.

You will need to replace the PLACEHOLDERS with your codes for splitting datasets (step 1), training (step 2), and testing (step 3). While there is no need to change the step 4, you are encouraged to implement your own ways.

In your report, please include both figures of sample scatters and quantitative results of your implementation.

Problem II. Confusion matrix

Suppose that there is a trained classifier for predicting the animal classes (e.g., cat, dog) of a photo. The following table lists the prediction class and ground-truth class for each test image.

Image ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
True class	C	C	C	C	C	D	D	D	D	D	D	D	D	M	M	M	M	M	M	M
Predicted class	D	C	D	D	M	D	D	C	C	M	M	D	C	C	C	M	M	D	D	M

Notes: C, cat; D, dog; M, monkey

Please manually compute and report the confusion matrix and accuracy. For each of the three categories, calculate its precision and recall rates.

Problem III. Comparative Studies

Please write a function to calculate the confusion matrix for the prediction results of a classifier. This function should take the form:

```
def func_calConfusionMatrix(predY, trueY)
```

where predY is the vector of the predicted labels and trueY is the vector of true labels. This function should return accuracy, per-class precision, and per-class recall rate.

Please use above function in the script “main_part1.py” and report the confusion matrix of both logistic regression implementations.

Please include in your write-up all numerical results and graphical results.

How to submit

- Submit your source codes and report through the **course site**. The codes should be self-contained, and run without any error. Otherwise, severe penalty will be applied.
- ***No hard copy required.***