

CS 436/536 Assignment-2

Description:

1. Get the data: Get the data and data description from the UCI repository about characteristics for 3 different flowers: <https://archive.ics.uci.edu/ml/datasets/iris>
2. Data contains 4 specifications:
 - sepal length in cm
 - sepal width in cm
 - petal length in cm
 - petal width in cm
3. It has three types of outputs as flower names: "Iris-setosa", "Iris-versicolor" and "Iris-virginica".
4. Make a logistic regression model using two features (for example: "sepal length" and "sepal width") as input and any two flowers (for example "Iris-setosa" and "Iris-virginica") as binary classifiable output.
5. The repository has 4(3 input and 1 output) features and 150 rows, as per instructions in step 4 the data you select will be 3(2 input and 1 output) features and 100 rows.
6. Plot the input and output data you selected for your model, clearly showing the data distribution.
7. Split your data in 80:20 (Train : Test) form, train your model with 80 Train data and plot the accuracy percentage on train data vs number of epochs.
8. After the model is trained, test the model with test data and report the accuracy percentage with test data.
9. Feel free to use any value of weights and any learning rate, report your weights after the training.

Submission:

1. Due date is **midnight of 14 March 2023**. The total points will be reduced by 5% for each day after the due date.
2. Submission in a single PDF file. Include all the code with comments, plots, and summary.
3. Segregate the data of your choice and plot its distribution. [10 Points]
4. Split the data randomly in 80:20, that is train and test data both should have roughly 50-50% data of each class. Print to show the split. [10 Points]
5. Build the logistic regression model and plot the curve of accuracy and epochs for train data. [60 Points]
6. Test your model and report the accuracy of test data. [20 Points]

Note:

1. It is important that everyone submits their entire notebook code in PDF format as well merged to the report. If this is not done points will be deducted.
2. Feel free to provide links to your notebook in your report as well.
3. Review and follow these [Watson College Academic Honesty policies](#) that spell out the consequences of academic dishonesty.
4. Do not copy/give code from/to others. If plagiarism is found, both will receive zero points.
5. You can submit multiple times before the due date, only the last submission will be graded.

Important Notes:

Make a report first answering all the questions individually. For example, if a question is asking you to show a plot, copy your plot from your notebook and paste it in your report with the question. If a question is asking you to report the final accuracy, please put the values in the report with the question.

After the report is done, then merge a pdf of your notebook to the bottom of your report. Please submit only a single PDF file.

Hint:

If your model is not performing well for the flowers/features you selected, try to change your selections.