



COMP338: Artificial Intelligence

Fall 2023/2023

Project Machine Learning: Regression

Deadline: 26 January 2024 (before 11:00PM)

Dr. Mohammed Helal & Dr. Radi Jarrar

In project you will build a predictive model based on Linear Regression algorithm. This is a team-work of groups (**each group of two students**). You should submit the code of the work as well a report containing a proper cover page and the details of your work.

The dataset will be used to construct the regression models is provided under this link:

https://www.dropbox.com/s/8ezou3ayuoehwl/Height_Weight.csv?dl=0.

Your task is to predict the weight of a person given their heights. The dataset contains weight/height instances of previously collected data that you will use to build your model. You are required to use the Weka Machine Learning library. You can use either Python or Java to import Weka library¹.

Your solution should contain the following:

1. In your code, convert the height from inches to cms and the weight from pounds to kilograms.
2. Print the main statistics of the features (i.e., mean, median, standard deviation, min, and max values) in a proper table.
3. For each of the following models, you have to split the data into 70% training and 30% test.
4. Select a subset of 100 instances from randomly selected from the dataset and generate the first model (called M1) and test this models performance using appropriate regression metrics.
5. Select a subset of 1000 instances from randomly selected from the dataset and generate the first model (called M2) and test this models performance using appropriate regression metrics.
6. Select a subset of 5000 instances from randomly selected from the dataset and generate the first model (called M3) and test this models performance using appropriate regression metrics.

¹ <https://waikato.github.io/weka-wiki/use-weka-in-your-java-code/>

7. Use the entire dataset and generate the first model (called M4) and test this models performance using appropriate regression metrics.
- 8. Print the appropriate performance metrics for Regression and compare the performance of the generated models in your own words.

You should have all points written properly and explained in the report. The report should contain your explanation of the work, details of your implementation and the results.

You have to turn in a softcopy of your code and a Word document containing the information required to as specified above. Please response to the thread where this assignment was sent with your solution. This is a group project and you should specify the file name as St1Name_St2Name_AIPML_YourSecNo.zip