

[Assignment 2](#)

[How to submit](#)

[Tasks](#)

[Notes:](#)

[What to submit](#)

[Important Links](#)

Assignment 2

Complete the assignment individually. If you worked in a group for assignment 1, you can use the same dataset that you worked with. But this time, you have to work alone.

How to submit

1. Create a folder, named like this: "lastname_firstname"
2. In the folder, include the report. The report has to be a single pdf file. **I will not accept any other file format.** The naming convention is like this: firstname_lastname.pdf. **Anyone who does not abide by this convention rule will get a penalty of 10 points.**
3. Include **JUST** the coding files (just the coding file and absolutely nothing else (that means .py, .rmd, .r, .ipynb etc). Any file/folder extraneous and 10 points will be the penalty for that.
4. Penalty for late submission: Every day 10 points.
5. Deadline: February 27, 11:59 PM

Tasks

1. Use the data that you have used in your assignment 1 and use Min-Max Scaling and Z-score scaling on it to transform your data (10 + 10)
2. After the scaling, use PCA and forward feature selection and backward feature selection to select the best 5 features for your data (10 + 10 + 10)

3. Any other scaling or feature selection that you find interesting, implement them and see what result you get. Try at least one other Scalar and one other Feature selection. See the links/references below for further exploration (10 + 10)
4. Write a report based on all your findings. (30)

Notes:

1. For this assignment, only work with numeric data
2. If you don't have at least 3 numeric attributes, you can work with this dataset.
 - a <https://www.kaggle.com/kumargh/pima-indians-diabetes-csv>
 - b Description of the data :
<https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.names>

What to submit

Only a .pdf file of your report. Include

the following in your report:

1. Write about what scaling methods you used and what results they gave. Are they different? How different are they? Include a screenshot of the results as proof. (10)
2. Describe the feature selection techniques that you used. How different are they from each other? How consistent are the results? Include a screenshot of the results as proof. (10)
3. If you do not use the scaling methods, how different do the results become for step 2? Include a screenshot of the results as proof. (10)

Important Links

- 1 <https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.names>
- 2 <https://scikit-learn.org/stable/modules/preprocessing.html>
- 3 https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html
- 4 https://scikit-learn.org/stable/modules/feature_selection.html
- 5 https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- 6 https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html#sklearn.feature_selection.RFE
- 7 **Please check out the video on BB:- link where i have a demo recorded.**