

Week 2 – SVM + PCA (formative assignment)

In this assignment you will perform an analysis with different models and hyperparameters and compare the outcomes.

As part of the assignment you will create, as specified in the rubric (criteria 1 through 3):

1. *a written justification of the choice for one or more machine learning models suitable for the problem.*
2. *a detailed analysis using an existing implementation.*
3. *a written evaluation of the model's quality using an appropriate criterion.*

Work out your answers to the assignments/questions below in a Jupyter Notebook. Clearly separate natural text in Markdown cells (used for explanations, justifications, conclusions and reflections) from program code (used for analysis or visualisations).

At the same time ensure the notebook still provides a cohesive account of the analysis.

Assignment Description

1. Find two datasets online; one with two features, one with three features (use PCA if necessary) with a binary label. Select these two datasets such that one is easily linearly separable, and one is not linearly separable.
2. *Do not submit the datasets as part of your deliverables, but refer to the location where it can be downloaded (URL).*
3. Give a clear description of the features of the datasets.
4. Train your support vector machine classifier with a linear kernel on the first group of data, and with a non-linear kernel on the second group. Experiment with some options for 'C' and 'gamma' of the SVM classifier and show their effect. *(criterion 2)*
5. Report on the quality of your models using appropriate evaluation metrics. Some suggestions have been made in the lecture, but don't hesitate to use other criteria with a suitable substantiation. *(criterion 1 & 3)*
6. For the second dataset only:
 - a. Visualise the decision boundary of the support vector machine classifier. *(criterion 2)*
 - b. Visualise a couple of support vectors of the support vector machine classifier. *(criterion 2)*

Assignment deliverables

Please submit the following three components:

1. Documented implementation

Submit a compressed (zipped) folder containing your Python scripts or Jupyter notebooks.

- Do not include generated output files or your virtual environment.
- If you used a virtual environment, include a requirements.txt file listing all required libraries.

Format: .zip

2. Printable Version of Your Code

Provide a PDF version of your code (either notebooks or stand-alone scripts). This helps us offer detailed, line-by-line feedback.

- a. Ensure code readability: use landscape orientation if needed to accommodate longer lines.

Format: .pdf

3. Individual Contribution & AI Usage Report

Write a short document addressing the following points:

- a. **Individual Contribution:** Describe which parts of the assignment you completed and what responsibilities you took on.
- b. **AI Usage:** Specify whether you used AI tools and, if so, how. Be transparent about the extent and purpose of any AI support.

Format: .pdf

Important Notes

1. Before submitting your notebook, restart the kernel and run all cells to ensure it executes cleanly from top to bottom. Remove any error messages or irrelevant outputs.
2. As with all your written assignments, you may build upon external ideas (including those from chatbots or AI tools), provided you cite your sources. Use APA or IEEE referencing style.
3. AI usage must be clearly documented. You may use AI tools as a source of feedback or inspiration, but the final work **must** be your own.