EAAA Erhvervsakademi Aarhus

Machine Learning

ML Mandatory exam

Peter Bundgaard Immersen

# Reflection

Exercise 1: Machine Learning Assignment with Various Models

My experience working on Exercise 1:   
Since this assignment was open for trying lots of stuff I found it very useful since it was both challenging my current knowledge and allowing me to refresh many of the areas we’ve had in this course.  
It was both challenging and rewarding. Most of my time on this assignment involved making different machine learning models on the make moons dataset to see which could classify it best.  
That meant going through the models mentioned below:

1. SVM: Support Vector Machines

2. Neural Net

3. Random Forest

4. Logistic Regression

5. K-means

6. Decision Tree

Exercise 2: Titanic Dataset with Neural Net and Random Forest

Exercise 2 focused on the Titanic dataset, where I trained both a neural network and a random forest model. Working with real-world data provided a different perspective compared to the synthetic data in exercise 1.

It was also a fun perspective having to search the web for which class of people embarked from the 3 different ports, and having that as a factor play in when training the neural network.  
Compared to earlier in this course where we didn’t have as many columns to train the model with, where the accuracy was 73%, now we are getting 80-84%

The excercise required feature scaling, handling categorical variables, and dealing with class imbalance. On the other hand, Random Forest, being an ensemble method, proved robust and required less parameter tuning, and with a max\_depth of 5 it also provided a very good result

# Exercise 3: Iris Dataset with PCA and Clustering Algorithms

Exercise 3 helped me getting a better understanding of dimensionality reduction techniques like PCA (Principal Component Analysis) and various clustering algorithms, including K-means and DBSCAN, using the Iris dataset. PCA helped visualize high-dimensional data in a reduced space, making it easier to understand the underlying patterns.

K-means and DBSCAN clustering offered insights into grouping similar data points together. It was interesting to observe how the choice of clustering algorithm and hyperparameters influenced the resulting clusters.

Where K-means actually did a way better job than DBSCAN which was also to be expected.

The final exercise involved image recognition using PCA and SVM on the Olivetti Face dataset. This task was very fun, as it dealt with image data. PCA was used for dimensionality reduction, preparing it to machine learning algorithms, SVM in this case, to image data.

Overall, this assignment had a significantly positive impact. You were able to write your own machine learning models instead of looking at code to see what it does, not that looking at code doesn’t also provide you with knowledge, for me it works better getting the code into your fingers.

By doing so, you had to work more independently, and it became necessary for you to debug the code you had written if it failed. This experience greatly contributed to a deeper understanding of machine learning models and their various use cases, highlighting the specific strengths of each model.