

Project 2: Image Classification



Problem Specification

The objective of this assessment is to build a machine learning Cat/Dog image classifier. You'll implement a system to classify whether images contain either a dog or a cat. This is easy for humans, dogs, and cats. Your classifier will find it a bit more difficult.



The dataset we will be using is the Dog Vs Cats. You can find the original dataset [here](https://www.kaggle.com/c/dogs-vs-cats/data).
<https://www.kaggle.com/c/dogs-vs-cats/data>

The dataset is comprised of 25,000 images of dogs and cats. This is a binary classification problem. The images have significant variation in pose.

Your goal is to build a machine learning algorithm capable of detecting the correct animal (cat or dog) in new unseen images. Below you can see a selection of images from the dataset.



Figure 1 Sample Of Cats and dogs images from Kaggle Dataset

You can use the full original dataset (from Kaggle) or you can find a selection of images (1000 for the train and 100 for the test) for an initial development in the file (**project3_files.zip**) in the assignment unit on Canvas. Please make sure that all your images are of an equal size (eg. 350 *350 *3) and the data has also been normalized.

The assignment consists of the following three sections:

- **Part A:** Pre-processing Phase:

In this phase you will be focussing on preparing the dataset, i.e. loading, displaying, normalisation, resizing, dimension reduction (PCA), etc.

[35 Marks]

- **Part B:** Training phase:

In this phase, you will train different machine learning algorithms using the pre-processed dataset comprised of the images and their corresponding labels split in train and development (a part of your training data you will use during the implementation). The test dataset will be used to predict labels of unseen images.

[40Marks]

- **Part C:** Optimization phase:

Focus on hyperparameter optimization on the final model.

[25 Marks]

Submission

Please upload your final submission (as a single zip file) to Canvas before **22:00 on April 10th**.

Your submitted zip file should contain your python and a report. Please note that all code should be fully commented. **Please do not include the dataset or models in your uploaded zip file.**

Code

All code should be completed using Python as the programming language. You can use Scikit Learn, NumPy and Pandas, OpenCV and Pillow. You are free to use imported graphical libraries such as [Matplotlib](#) or [Seaborn](#).

Your code should have a logical structure and a high level of readability and clarity. Please comment your code and put all code into functions. Your code should be efficient and should avoid duplication.

Late submissions

If you don't get the assignments done to your satisfaction and don't meet the minimum requirements by the deadline, you have the option (as with any assignment at MTU) of submitting up to 1 week late for a penalty of 10%.

This penalty is subtractive. Work that would have earned 55% if on time, would get 45% (not 49.5%) if late.

The penalty is applied weekly. So 1 day late costs the same 6. If you want to take that option please let me know. Otherwise, I will just correct whatever I have.

If you have a specific reason for submitting a late assignment (sickness, etc) please contact me directly or submit a medical certificate in the department secretary.

Plagiarism

Please read and strictly adhere to the [MTU Honesty, Plagiarism and Infringements Policy Related to Examinations and Assessments](#). Note that reports are **checked** against each other and against external web sources for plagiarism. Any suspected plagiarism will be treated seriously and may result in penalties and a zero grade.

Grading

The assignment is worth 40% of the overall mark for the module. Marks will be awarded based on the quality of the code and the results. In particular, I will be checking to see if you are handling data correctly, carrying out exploratory analysis to gain insights, correctly performing model selection, and critically, documenting everything in a clear and concise way. The submitted code will also be checked to ensure that the work is your own.