Coding Homework 1

> Krishna Penukonda
>
> Student ID: 1001781

Instructions to run

Split the dataset

`split.py` is responsible for extracting the relevant samples from the full dataset and splitting them into train, val and test. This script requires the concepts file, unzipped ImageClef features directory, and the annotations file. By default, they need to be in the same directory as `split.py`

```
./split.py
```

The split dataset is placed (by default) in `dataset/`

Train classifier

```
./train.py
```

Trains and evaluates a (on-vs-rest, SVM-based) classifier.

Requires that `dataset/` be present and that it contains `x_train.npy`, `y_train.npy`, `x_val.npy`, `y_val.npy`, `x_test.npy`, `y_test.npy`, and `classes.txt`

Class-wise dataset splitting

Splitting each class into train, val, and test sets individually as opposed to splitting the entire dataset (or a subset consisting of multiple classes) ensures that the ratio of positive to negative samples in each class remains the same across the splits.

Results