**Traffic Sign Data Set:**

This data set is subset the computer vision benchmark data set “*German Street Sign Recognition Benchmark by* Stallkamp et al”. The original data set consists of 39,209 RGB-coloured train and 12,630 RGB-coloured test images of different sizes displaying 43 different types of German traffic signs. These images are not centred and are taken during different times of the day.

You will be working with a sample of this data set which consists of 10 classes and 9690 images. The images have been converted to grey-scale with pixel values ranging from 0 to 255 and were rescaled to a common size of 48\*48 pixels. Hence, each row (= feature vector) in the data set has 2305 features and represents a single image in row-vector format (2304 features) plus its associated class label.

We changed the class labels from the original dataset so the classes we use are now labelled from 0 to 9 indicating the following traffic sign categories:

|  |  |
| --- | --- |
| **Class label** | **Traffic Sign** |
| 0 | **speed limit 20** |
| 1 | **speed limit 30** |
| 2 | **speed limit 50** |
| 3 | **speed limit 60** |
| 4 | **speed limit 70** |
| 5 | **left turn** |
| 6 | **right turn** |
| 7 | **beware pedestrian crossing** |
| 8 | **beware children** |
| 9 | **beware cycle route ahead** |

Compensating the light conditions and position of the images is not necessary for the coursework, i.e. you are not expected to apply complicated image processing techniques.

Below are examples of images of the street signs in this data set:

**A group of images of signs

Description automatically generated**

You can access the data **.csv files** by following the link [Traffic\_Sign](https://heriotwatt.sharepoint.com/sites/Neamat_F21DLMaretial_2021/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FNeamat%5FF21DLMaretial%5F2021%2FShared%20Documents%2FCW%5FDataSets%2FTraffic%5FSign&p=true&fromShare=true&ga=1)

The .csv files labelled with *x\_train* contain the image data for each sign and the .csv files labelled with *y\_train* contain the corresponding class label for each sign in the same order. This is a standard for data representation in machine learning. The test data sets are provided similarly as *x\_test* and *y\_test*. Note you might not need to use the test data in your data exploration steps.

You will see that we provide 12 different versions for train/test splits indicating the below:

1. Complete Dataset: { x\_train\_all/y\_train\_all and x\_test\_all/y\_test\_all }
2. Binary Classification Data set[[1]](#footnote-1): This dataset has only two classes (class 0 and class 1 – for binary classification). { x\_train\_bin/y\_train\_bin and x\_test\_bin/y\_test\_bin }
3. 10 one-vs.-rest samples in folder OnevrsAll[[2]](#footnote-2): The training pair **[x\_train\_all.csv]** and **[y\_train\_<label>.csv]** contain training features and labels for *one-vs-rest classification*. In each file, the images with class <label> have a 0 and all the other images a 1. For example, if the <label> is 6 (as in **y\_train\_6.csv**) then all the images of right turn signs have a 0 as their label and all the other images have a 1. { using these 10 data sets can help you examine the performance related to each class separately}

1. This is a reduced data set and would be less computationally intensive to run. Using this data set can help you when you want to run preliminary experiments or when you reach computational limits. [↑](#footnote-ref-1)
2. ‘one-vs-rest’ data sets can be used as reserve data sets, for testing various hypotheses you may come up with in the coursework. Also they may give better accuracies, and thus may be handy for some experiments. Please use them to enrich your research hypotheses and experiments. [↑](#footnote-ref-2)