# CIFAR 10 task

## Data Preparation

This part was straightforward: following the directions on the assignment, the data was loaded, normalized and flattened. Then, 80% of the training data was dedicated to training and the rest to validation.

## Problem Analysis

A quick look at the dataset, including visualization of samples, revealed that the classes look well-balanced, with exactly 5.000 samples per class. The problem at hand is undoubtedly a multi-class classification task, for which accuracy was chosen to be the target.

## Model Selection and Evaluation

The three models chosen for this task were SVM, Gaussian Naıve Bayes and ANN. As a bonus test, it was also decided to add both the linear kerneled Support Vector Machine and the Gaussian kerneled Support Vector Machine. The results of the experiment are:

|  |  |
| --- | --- |
| Models | Accuracy (%) |
| ANN | 44.48 |
| Linear SVM | 34.46 |
| RBF SVM | 49.45 |
| Gaussian Naıve Bayes | 29.39 |

These are the results of the most successful experiment. As can clearly be seen, the best model was the Gaussian kernel SVM, followed by the feed-forward model, then by the linear kernel SVM, with the Gaussian Na¨ıve in last place. The training time for these models are:

|  |  |
| --- | --- |
| Models | Training time (s) |
| ANN | X |
| Linear SVM | 1184 |
| RBF SVM | 1066 |
| Gaussian Naıve Bayes | 3 |

The Naıve Bayes could be trained in just a handful of seconds, while the two SVM took slightly more than a quarter of an hour to finish. The feed-forward model took several hours, the reason being that a very high number of epochs was used. Arguably, an early-stopping mechanism could have been used, since the validation accuracy and loss displayed the tendency to plateau before the first 10 epochs, but due to the general lack of improvement and the overall well- established results, the group decided to move on and focus the efforts elsewhere. Nonetheless, the feed-forward still proved to be the slowest to train, even with a reduced number of epochs. No significant differences in how models predict categories was found in the interpretation of results.

## Cross Validation

As per homework assignment, stratified K-fold cross validation was used. Here are the results for k=10:

|  |  |  |
| --- | --- | --- |
| Models | Mean accuracy (%) | Standard deviation (%) |
| ANN | 41.83 | 1.25 |
| Linear SVM | 33.49 | 1.15 |
| RBF SVM | 48.83 | 1.51 |
| Gaussian Naıve Bayes | 28.00 | 0.84 |

The results are in line with the test accuracy, as the mean roughly reflects the outcome of the best experiment reported in the previous section. The most consistent across folds was definitely the Naıve Bayes, as can be seen from the relatively low standard deviation. Typically, cross-validation is more representative of a model’s performance than a train/test split, because it takes into account the average of multiple experiments, ensuring that the results are as free as possible from noise (such as a coincidentally favorable partition of the test set).