

DATA.ML.100 Introduction to Pattern Recognition and Machine Learning
TAU Computing Sciences
Exercise 2 *Visual classification (CIFAR-10 dataset)*

Be prepared for the exercise sessions (watch the demo lecture). You may ask TAs to help if you cannot make your program to work, but don't expect them to show you how to start from the scratch.

1. **CIFAR-10 – get to know** (10 points)

CIFAR-10 is a benchmark popular datasets used in machine learning literature for bechmarking different methods. For brief introduction check <https://www.cs.toronto.edu/~kriz/cifar.html>.

Download the dataset to your local computer (perhaps your prefer `/worktemp` instead of your home directory) from https://tuni-my.sharepoint.com/:u:/g/personal/joni_kamarainen_tuni_fi/EdgoMFSFCKRGiM3QcSZtQ9VISA?e=Ye4vs8. Extract the files and then edit configuration paths in the provided `cifar10_illustrate.py`. Run the code:

```
(intro_prml) kamarain@Joni-Precision-5520:~/exercises$ python cifar10_illustrate.py
```

The code should randomly plot images from the dataset. Note that the code loads only one of the training batches that each contains 10,000 training samples. Test data is also provided separately (10,000 samples, 1,000 per each class).

2. **CIFAR-10 – Evaluation** (10 points)

Write a python function `class_acc(pred,gt)` that computes the classification accuracy for predicted labels `pred` as compared to the ground truth labels `gt` (`Y` values in the provided code). Make sure that your function works properly by using the true labels of CIFAR-10 test samples that should provide perfect accuracy.

3. **CIFAR-10 – Random classifier** (20 points)

Write a function `cifar10_classifier_random(x)` that returns a random class label for the input data `x`. Make a script that inputs all CIFAR-10 test samples to this function and evaluates its classification accuracy using your evaluation function.

4. **CIFAR-10 – 1-NN classifier** (20 points)

Write a function `cifar10_classifier_1nn(x,trdata,trlabels)` that finds the best match of the input vector `x` in the training set `trdata` and returns the same label. Make a script that inputs all CIFAR-10 test samples to this function and evaluate its classification accuracy using your evaluation function.

Make sure your 1NN function works by testing with the training data itself (should provide perfect accuracy).

Hints: During developing stage you might want to use only a sub-set of the full data to make evaluation faster.