

SGN-13006 Introduction to Pattern Recognition and Machine Learning
TAU Computing Sciences
Exercise 2 *Visual classification (CIFAR-10 dataset)*

Be prepared for the exercise sessions. You may ask TA questions regarding your solutions, but don't expect them to show you how to start from the scratch. Before the end of the session, demonstrate your solution to TA to receive exercise points.

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

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

Download the dataset to your local computer (perhaps you prefer `/worktemp` instead of your home directory) from https://tutfi-my.sharepoint.com/personal/joni_kamarainen_tut_fi/_layouts/15/guestaccess.aspx?docid=04c0040d0ce9e4bf3993b2a847fdf931f&authkey=ATeq2eP1RwVY8IQuZdjyBI8. Extract the files and then edit configuration paths in the provided `cifar_10_read_data.m` Matlab script and make sure you can run it!

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

Write a function `cifar_10_evaluate(pred,gt)` that computes the classification accuracy for predicted labels `pred` as compared to the ground truth labels `gt`. 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 `cifar_10_rand(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 `cifar_10_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.