# SOI1010 Machine Learning II - Assignment #2

Due: November 8, 2023 11:59 pm

The submission should include a code (both link to the colab and .py format) and a report that has answers to the questions and results. Use ***PyTorch*** (or TensorFlow/JAX). Also, minimize the use of numpy. There will be mark deductions if numpy is used when PyTorch is preferable or should be used. Marks will be deducted if the submission does not include the requested files. **DO NOT** use other libraries, such as scikit-learn/sklearn, to use a model (kNN in this case) you are supposed to implement. **Using sklearn or any other third library or already built-in functions** that you are asked to implement **will result in 0 mark**. So, please don't ask "Can I use this library/numpy/etc.?" Also, if an assignment asks you to implement some model, that means you shouldn't use the built-in implementation from any library for that model in the first place.

## Problem #1: Binary Classification via soft-margin SVM on CIFAR10

a) Load CIFAR10 dataset as follows:

```
import numpy as np
import torch
from torchvision import transforms, datasets
trainset = datasets.CIFAR10(root='./data', train=True,
                                        download=True,
                                        transform=transforms.ToTensor())


testset = datasets.CIFAR10(root='./data', train=False,
                                       download=True,
                                       transform=transforms.ToTensor())
```

b) Visualize at least one image for each class. You may need to look into how dataset is implemented in PyTorch.
c) Split the trainset into training set and validation set with 90% : 10% ratio. Implement

dataloaders for CIFAR10.
d) Choose any two classes. Then, make a SVM classifier (implement a loss function yourself. Do not use PyTorch implementations of loss functions.) and its training/validation/evaluation code to perform binary classification between those two classes.
e) Train for 10 epochs with batch size 64.
f) Perform data normalization. You may need to look into how to use datasets in PyTorch.
g) Again, train for 10 epochs with batch size 64 after data normalization. Write down your observations.
h) What are the hyperparameters you can tune?
i) Try to obtain find optimal hyperparameters.
j) What is the final *test* accuracy?

## Problem #2 [Bonus/Optional]: Multiclass Classification via soft-margin SVM on CIFAR10

a) Perform multiclass classification using soft-margin SVM on the whole dataset.
b) Perform hyperparameter search.
c) What is the final *test* accuracy?