This excersise is accompanied by the `assignment_3.zip` archive in this folder. It holds 3 files that you will need for this assignment: `samples.npy`, `model1.hdf5`, and `model2.hdf5`.

You can load the models to python with `model1 = keras.models.load("model1.hdf5")` and the samples with `samples = numpy.load("samples.npy")`.

A

18 points

`samples.npy` holds 100 samples. 50 of it were part of training set `1` of `model1.hdf5` and 50 were part of training set `2` of `model2.hdf5`.

- For each sample, identfiy which training set this sample was part of. Note, you probably will not be 100% certain - an educated guess is fine.
- If you cannot make an educated guess for a sample explain why. It may not be possible to identify each sample!
- Explain and argue, why you think samples are part of which training set.
- Please provide a list with 100 labels ('1' / '2') indicating the training set you think the respective label was part of. Please use the same order than `samples.npy`.

You are free to use any technique to help you in your identification. Please indicate what exactly you did and explain your thinking. You are free to use any information available to you. This includes:

- self trained models from previous excercises
- models you found online
- the MNIST dataset including labels

B

7 points

Both training sets `1` and `2` were unbalanced.

- What can you infer about the training set's imbalances?

You are free to use any technique to help you in your identification. Please indicate what exactly you did and explain your thinking. You are free to use any information available to you. This includes:

- self trained models from previous excercises
- models you found online
- the MNIST dataset including labels

C

5 points

- Describe challenges you faced in A and B and how you solved it.