AI1072: Machine learning, exercise sheet 2

MNIST can be obtained via

```
wget www.gepperth.net/alexander/downloads/mnist.npz
```

Exercises

a) Load the MNIST data as follows:

```
traind,testd,trainl,testl= np.load(open("mnist.npz","rb")).values() ;
```

'traind' contains the samples, 'trainl' the target values (labels). Inspect both arrays and especially print out their shape. How many samples and targets are there? How many numbers constitute a sample? Why are there 10 dimensions to a single target value?
b) Slice out the 1000th sample. What is its class?
c) Calculate the lowest and highest class in the labels!
d) Calculate the number of samples having class 5 (tip: generate a boolean mask and uset he fact that a bool in numpy is in fact an int with value 1, whereas false has value 0)
e) Calculate the smallest and highest pixel value in the 10th data sample
f) Generate the following variations of the 10th sample and display them successively:
- just keep every 2nd row
- just keep every 2nd column
- inverse all rows and all columns
- set the 10 topmost columns to 0
- set the 10 lowermost columns to 0
- invert rows, invert colums, just take every 2th row and every 2th column

g) Extract from the training data just the samples having class 4, same goes for the labels.
h) Extract from the training data just the samples having class 4 or 9, same goes for the labels.
i) Extract the first 10000 data samples and targets!
j) Apply the in-place transform

$$1 - x$$

to all samples.
h) Copy out 1000 randomly chosen samples.