LCPB 21-22 exercise 1 (Gradient descent & Deep Neural Networks, DNNs)

In addition to the code written during the lesson, for the grid-search part (see below), consider notebook NB11 by Mehta et al., which is on GitHub or this website:

http://physics.bu.edu/~pankajm/MLnotebooks.html

One member of the group will submit the Jupyter notebook compiled, with all figures, by using the interface in moodle.

The name of the notebook should be ***group23??_exercise1.ipynb*** *where ?? is the code of the group.*

The notebook should start with the list of students (Name, Family Name, matriculation number).

Analyze the labeled 2d data x=(x1,x2) introduced in the second lesson, with DNNs.

1.
Implement a "grid search," as shown in NB11, to improve at least two of the aspects or hyperparameters of the model. One aspect is the choice of the minimization algorithm (ADAM, RMSprop, Nesterov, etc.). In the grid search, put at least a second aspect picked from this list:

- activation units (sigmoid, ReLU, ELU, etc.)
- dropout values
- learning rate
- size or number of layers
- rescaling of data

2.
Take the best DNN from point 1 and study how its performance changes when the number N of samples is

a) reduced
b) increased
c) *augmented*

Point 2.c means taking the given N=4000 samples, splitting them in training and validation, and "augmenting" the training samples by generating artificial ones similar to the real ones. For instance, a given training sample x=(x1,x2) can be transformed to multiple copies (x1+s1, x2+s2) where each (s1,s2) is a *reasonably* slight random shift. Explain your choice for the variability of this shift. Why do not we also augment the validation samples?

OPTIONAL
3.
Explore what happens with data generated by the second nonlinear function or other choices.