

# Project 3

(M1) is a neural network with topology and activation function of your choice (differentiable or not), but mandatory  $L_1$  regularization.

(M2) is a standard  $L_2$  linear regression (least squares).

(A1) is a standard momentum descent approach

[reference: <http://www.cs.toronto.edu/~fritz/absps/momentum.pdf>] applied to (M1).

(A2) is an algorithm of the class of accelerated gradient methods

[references: <https://www.cs.cmu.edu/~ggordon/10725-F12/slides/09-acceleration.pdf>, <http://www.cs.toronto.edu/~adeandrade/assets/aconntmftc.pdf>, <https://arxiv.org/pdf/1412.6980.pdf>] applied to (M1).

(A3) is a basic version of the direct linear least squares solver of your choice (normal equations, QR, or SVD) applied to (M2).

No off-the-shelf solvers allowed.