To-do

Part 1 #1 - Cecilia

Part 1 #2 - Daniel

Part 1 #3 - Ilyes

To do on part 1)

* Tune hyperparameters for regular batch SGD, mini-batch, and momentum
* Implementation and graphical representation of the training/validation accuracy/error for the different methods
  + (e.g. training and validation accuracy as a function of number of training iterations/value of learning rate)
* Implementation and graphical representation of convergence time for the different methods (can just use a basic function timer)
* Use the finalized models on the test sets.
* number 4) of part 1, aka mini-batch and momentum together

Notes:

Daniel:  
sources:

<https://en.wikipedia.org/wiki/Stochastic_gradient_descent>

<https://www.geeksforgeeks.org/ml-mini-batch-gradient-descent-with-python/>

<https://ruder.io/optimizing-gradient-descent/index.html#minibatchgradientdescent>

<https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd/>