**Reproducibility Challenge**

ACCELERATING SGD WITH MOMENTUM FOR OVERPARAMETERIZED LEARNING

* **Claims/Goals – we need to prove or disprove this**
  + MaSS converges for the same step sizes as Nestrov SGD
  + MaSS obtains an accelerated convergence rates over SGD for any mini-batch size in the linear setting.
  + For full batch, the convergence rate of MaSS matches the well-known accelerated rate of the Nesterov’s method
  + Experimental evaluation of MaSS for several standard architectures of deep networks, including ResNet and convolutional networks, **shows improved performance over SGD, SGD+Nesterov and Adam**
* **Their steps**
  + **Synthetic data** 
    - Generate data using coupled decoupled model
    - Set random weights
    - Optimise MSE loss function
    - Test SGD + Nestrov. MaSS, SGD +HB, ASGD
  + **Real Data**
    - See Appendix H.1 and H.2 for their architectures and hyperparameters
    - MNIST testing with FCN
    - CIFAR-10 with CNN
    - Gaussian Kernel regression
    - Hyperparameter selection through dense grid search
    - Batch size 64
    - Momentum = 0.9
    - **Standard practices of data augmentation and reduction of learning rate – See section H2**
* **Tasks** 
  + Load CIFAR-10 and MNIST datasets
  + **Set up ReSNet (Get structure from another paper or code itself) - Conor**
  + **Set Up CovNet – Ahmad**
  + **Set up FCN - Ethan**
  + Build Nestrov optimiser -
  + Build Mass Optimiser
  + Test optimisers on networks
* **Preliminary tasks**
  + Get FCN, CNN working with ADAM and SGD
* **Other Tests we can do**
  + Hyperparameter search – See if dense grid search gives different hyperparameters
  + Test theoretical hyperparameters??