# Assignment Set 8

Pegah Khazaei 27th November 2020

---

## Different Optimizers

**Adam** is great, it's much faster than **SGD**, the default hyperparameters usually works fine, but it has its own pitfall too. Many accused **Adam** has convergence problems that often **SGD** + momentum can converge **better** with longer training time.

The **RMSprop** optimizer **is** similar to the gradient descent algorithm with **momentum**. The **RMSprop** optimizer restricts the oscillations in the vertical direction. **Adam** in some areas does not converge to an optimal solution, so for some tasks (such as image classification on popular CIFAR datasets) state-of-the-art results are still only achieved by applying **SGD** with momentum. **ADAM** is just **Adadelta** (which rescales gradients based on accumulated "second-order" information) plus momentum (which smooths gradients based on accumulated "first-order" information). I.e. **ADAM** is an extension of **Adadelta**, which reverts to **Adadelta** under certain settings of the hyperparameters.

# MNIST dataset