1. In continuation from the previous programming assignment on training the weights of a multilayer perceptron, implement each of the following gradient descent methods discussed in class. **Do not use built in functions**.

   (a) Stochastic gradient descent. (10)

   (b) Momentum. (10)

   (c) Nesterov Momentum. (5)

   (d) AdaGrad. (10)

   (e) RMSProp. (10)

   (f) Adam. (10)

   Clearly illustrate how the optimization proceeds in each case and summarize your observations. Use a dataset of your choice from the UCI data repository: click here.