COMP3222/6246 Coursework 2
Machine Learning with Python & Scikit-Learn

1. Motivated by the use of CNN for MNIST dataset, now you are to try CNN on the Sign Language MNIST next. In this task, you are given a dataset of hand gestures that are labelled with an American Sign Language letter (except for J and Z which need hand motion). The dataset can be retrieved from either `https://drive.google.com/file/d/1zkX8oQ74JFcJ7Gli6M_qnnGo41tMeZQI/view?usp=sharing` or the original source on Kaggle. Its format is very similar to the classic MNIST (28×28 pixels and each label is a number representing a letter from A to Z — but excluding 9=J and 25=Z). It is already partitioned into a training set and a testing set.

   (a) Implement a CNN and train it on the training set with the Gradient Descent optimiser. You are free to set the structure and the hyperparameters by your own. However, please write a short description of them and a justification of your choices.

   (b) Now replace the Gradient Descent with a Stochastic Gradient Descent (SGD) optimiser. Demonstrate how much does the CNN improve. Also justify your demonstration technique — why did you demonstrate in such a way? (You must also explain if there is any change to the CNN's structure).

   (c) Finally, replace SGD with the Adam optimiser and redo the previous subtask. Is it better to use Adam?

2. In this task, you have to do time series prediction using RNNs. In particular, we aim to predict the internet traffic data (the dataset itself can be found here - `https://drive.google.com/open?id=1DeOc_xdGFVboa7wOCNtcrrEx6LPBvIXM` - or you can find the link to this data set from the module web site). It contains internet traffic data (in bits) of a academic backbone network in the UK. It was collected between 19 November 2004 and 27 January 2005. Data were collected at five minute intervals.

   (a) Implement an RNN and train it on the provided data. You are free to set the hyperparameters by your own. However, please write a short description to justify your choices.

   (b) Once you have trained the RNN, use the Stochastic Gradient Descent technique to improve the RMSE and plot the results to demonstrate

how SGD works (again, you are free to choose the way you demonstrate - however, you need to explain why you did choose to do that way).

(c) Finally, replace SGD with Adam and redo the previous subtask. Is it better to use Adam? Explain your answer (hint: compare the number of steps, how many times it takes to converge etc).