Assignment 3

Submitted by,
Swati Verma
MT19073

Steps performed for classification of B-cell epitopes and non-epitopes using deep learning.

1. Initially for given training and testing set p-features are obtained by computing dipeptide sequences. Tripeptide sequences are also calculated but are not used further as they are not giving higher scores.
2. Calculated dipeptides for training and testing dataset are stored in a csv file called dipep_train.csv and dipep_test.csv.
3. Columns with zero values are removed.
4. A feed forward neural network has been applied by feeding 400 features as input to the perceptron.
5. Activation function used for input layer and hidden layer is hyperbolic tangent(tanh) and for output layer activation function used is sigmoid. Use of other activation functions were not giving a good score.
6. The model is then trained and fitted with different batch sizes and epochs values.
7. The two highest scores are obtained by setting batch size=100 and epochs =100 with the score of 0.78292. The output of this model is stored in a file named output_ANN.csv.
8. The score obtained by setting batch size=100 and epochs=150 is 0.76015. This output is stored in a file named final_output1.csv.
9. Both python file and notebook file are submitted.

Filenames given in folder are:

trainset.data: Training dataset downloaded from Kaggle.

testset.data: Testing dataset downloaded from Kaggle.

dipep_train.csv: Training dataset obtained by calculating dipeptide residue sequences

dipep_test.csv: Testing dataset obtained by calculating dipeptide residue sequences.

output_ANN.csv: Output file obtained by setting batch size=100 and epochs =100.

Final_output1.csv: Output file obtained by setting batch size=100 and epochs =150.

To run the python notebook file:

1. Open google colab in your browser
2. Open a new jupyter notebook and upload all the training and testing files given in the folder.
3. Upload the notebook file named BDMH_A3_ANN provided inside the folder.
4. Run one by one each cell until the output file is generated.
5. Download the output file generated and upload it to the submit prediction page of Kaggle link provided in the assignment page.
6. Submit prediction to see the obtained score.