Assignment-3 Project Report

- Convolutional Neural Network is used in classifying Transcription Factors Binding and Non binding sites
- Different Kernel sizes, filter sizes, Dropout values,learning rates, activation functions, optimization techniques are used and evaluated. The report presents the parameters with which I have got best results.
- Each genomic sequence is considered as an image of size 1x14 with 4 channels A,C,T,G.
- First, the data is imported and processed using Pandas. Each DNA sequence is hot encoded making them into a matrix of size 1x14x4.
- So treating the DNA sequence as an image a convolutional network is created with a filter size of 256 and kernel size of 1x24
- Two dimensional Max pooling layer, Fully connected Neural Network of units are added after the Convolutional Layer
- The output layer is made of one Neuron and Sigmoid activation function is used at this layer
- ReLU is used by default except in the output layer.
- A dropout of 0.5 is used to prevent Overfitting of the Network
- Binary Cross Entropy function is used as loss function and Stochastic Gradient Descent is used as the Optimizer.
- Early stopping is also implemented with patience 3 to avoid overfitting.
- Model is trained with a batch size of 200 and with 100 epochs.

Reference: Zeng, Haoyang et al. “Convolutional Neural Network Architectures for Predicting DNA–protein Binding.” *Bioinformatics* 32.12 (2016): i121–i127. *PMC*. Web. 9 Apr. 2018.