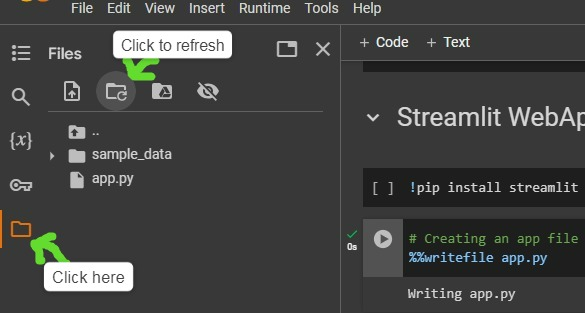
The given Google colab file aims to develop an LSTM-based model to predict the number of CpGs pairs given a DNA sequence. The code has two parts: Part 1 aims to solve the problem of count prediction for a fixed-size DNA sequence (128), and Part 2 aims to create a more generalized and size-free LSTM model that can take any length of DNA input (< 128) as input and still give an accurate prediction for the same.

Finally, a simple Streamlit-based web application is developed to showcase the performance of the model. The web application takes as input a string of DNA sequence and outputs the value of the number of CpGs pairs predicted by the LSTM model. The web application will also show the integer prediction made by the model and then compare it with the actual value.

To run the model and make the prediction, use the following steps:

1. Ensure GPU is available on the Google colab. It is needed for faster calculations.
2. Run Part 2 of the Google colab file. (Part 1 is not necessary, as Part 2 is an advanced and generalized version that works for input of all sizes).
3. Go to the “Streamlit WebApp” part of the colab file and run the first two cells of code. This will import Streamlit on the colab environment and create a file called “app.py”.
4. Go to the “app.py” file by double-clicking on it as shown in the below image and copy-paste the entire code present in the 3rd cell of the “Streamlit WebApp” part. If the file is not visible, refresh it by clicking on the icon shown in the image. (Uncomment the code)



1. On the Google colab notebook, run the next cell of code and copy the address generated.
2. Run the next cell of code and visit the URL generated. Paste the address copied earlier and hit “Submit”.
3. You will now be taken to the web application. Enter the DNA sequence you want to make the prediction for and click on “Make Prediction”.
4. Your predictions will now be visible on the screen.

As the model is trained on samples of length 64-128, to ensure high accuracy of prediction, make sure that the input given to the web application is within this range.

Try the model with different inputs. Thank you!!