HW5 Kaggle Competition Report

GitHub Link

Problem Description:

- Histopathologic Cancer Detection is a Kaggle Competition where the goal is to detect tumor cells located within the center of the images. The image size is 96x96 pixels while the tumor cell(s) would need to be located in the center 32x32 pixels part of the image.
- First, a dataframe is created from the dataset that is stored in the file train_submission.csv.
- From this, it is determined that there are 220025 images with two labels: 'id' and 'label', where 'id' is the input and it contains the file names and 'label' is the output that is 0 if there is no tumor or 1 if there is tumor. There are 130908 negative cases and 89117 positive cases. There are no missing/NaN values in the dataset.

Exploratory Data Analysis:

- A bar graph was plotted that displayed the number of negative and positive cases.
- Since there were no missing/NaN values in the dataset, no data was removed.
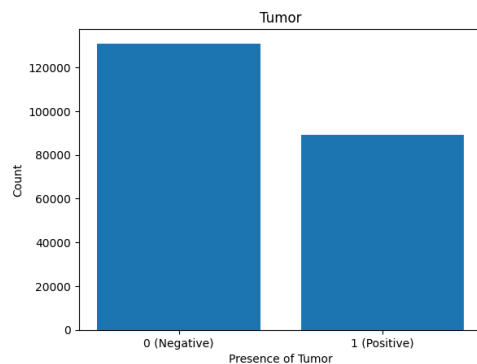
Model, Result, Conclusion:

- To build a model, I followed the instructions from the provided file (fashion-mnist-with-keras.ipynb):
  - First, I reformatted the labels.
  - Next, I built a model using model = tf.keras.Sequential() with two layers of activation relu and softmax, with optimization of SGD and learning rate of 0.1.
  - After compiling the model, I tried to train the model with model.fit(data_images, data_labels_, epochs=5) but received an error that stated that the input shape needed to be 9216 but it had received shape (None, 1). I had selected the input_dim to be 9216 because if the image was flattened then it's array would be 96*96=9216.
- I tried to research how to solve this error but was unable to find a solution.