Problem Set 1 (take at home - 40 points)

In what is described below we use the term dx to denote either a Gradient (vector) or a Jacobian (matrix) - it depends on the dimensions involved.

A. Convolution Operation Backpropagation (5 points)

You are given a matrix X of dimensions $n_x \times n_x$ and a kernel W of dimensions $n_k \times n_k$. You are also given a upstream dz of dimensions $n_z \times n_z$. Write the expression of the downstream dW when we have neither strides (s=1) or padding (p=0).

PS: If this helps, feel free to write the requiested calculation using specific dimensions of the variables involved.For example $n_x = 3, n_k = 2$.

B. Max-Pooling Layer Backpropagation (5 points)

You are given a matrix X of dimensions $n_x \times n_x$ and a kernel W of dimensions $n_k \times n_k$. You are also given a upstream dz of dimensions $n_z \times n_z$. Write the expression of the downstream dX when we have a max-pooling operation.

PS: If this helps, feel free to write the requested calculation using specific dimensions of the variables involved.

C. Architecture for binary classification (15 points)

Draw a CNN architecture that includes few CNN layers and is able to perform classification on the Fashion MNIST dataset: https://github.com/zalandoresearch/fashion-mnist. Please note that for this question, the URL of the dataset is given just for you to understand the input dimensions as well as the number of classes involved. Define the network to your liking - it can be as simple or as complex as you can make it but consider the performance requirement of step D below. The network will be considered correct if all dimensions involved are fully specified.

D. Implementation (15 points)

For are asked to code up **from scratch** a CNN network that implements the architecture you came up with in C above. In your code **you are asked to use only numpy / scipy and plotting libraries**. You also need to show clearly in markdown latex math notation, all the backpropagation equations involved. If repetitive layers / blocks are involved show the backprop equations for only one of the many.

PS: If it makes sense, you may use the results from A and B questions above for some of the backprop equations. To get full points the network must be trained, well documented in conjunction with your answer to C above and produce above 80% accuracy.

Submit your **Colab notebook URL** after sharing it with the grader and the professor to the URL that will be given to you **at the date of the exam**. No PDFs will be accepted as an answer to Problem Set 1 - all explanations must be inserted / typed in the notebook itself. If you handwrite anything please include it as an image in the notebook itself.

This is a exam problem. If you are found you communicated or collaborated with anyone to submit your answer, all parties involved will receive a grade of 0 for the whole midterm exam and will be referred to your department.