

### Question: Lost in the closet (Classification) .....

You are an artist who secluded yourself for years to come up with the perfect design for a new brand of clothes. However, your time off civilisation was not so beneficial since you cannot distinguish a T-shirt from a dress or a sneaker from a sandal anymore! In order to redress (!) that issue, you choose to train a Convolutional Neural Network (using PyTorch) that will help you identify each cloth to match the perfect design you created. In order to train it, you decide to rely on the dataset [fashion MNIST](#).

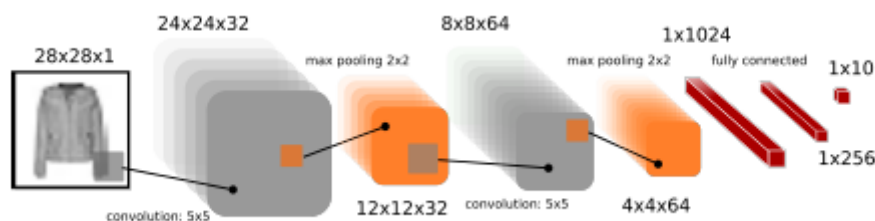
You can access the data using the following lines:

```
import torchvision
import torchvision.transforms as transforms
import torch

train_set = torchvision.datasets.FashionMNIST(root = ".", train = True , download = True , transform =
transforms.ToTensor())
test_set = torchvision.datasets.FashionMNIST(root = ".", train = False , download = True , transform =
transforms.ToTensor())
training_loader = torch.utils.data.DataLoader(train_set , batch_size = 32, shuffle = False)
test_loader = torch.utils.data.DataLoader(test_set , batch_size = 32, shuffle = False)
torch.manual_seed(0)
# If you are using CuDNN , otherwise you can just ignore
torch.backends.cudnn.deterministic = True
torch.backends.cudnn.benchmark = False
```

(a) Given the problem, what is the most appropriate loss function to use?

(b) Create and train a Convolutional Neural Network corresponding to the following architecture:



For training, initialise your weights using the Xavier initialisation, use ReLU as the activation function, a learning rate of 0.1 with the SGD optimiser. You will train your neural network over 50 epochs. What is the final (train and test) accuracy obtained? Provide a plot with the accuracy on the training and test set per each epoch. Looking at the loss through the epochs, discuss what you observe.

(c) Now, change the activation function to Tanh, Sigmoid and ELU. Provide only the final classification accuracy. Keeping ReLU, use 5 different learning rates: 0.001, 0.1, 0.5, 1, 10. What do you observe? Explain.

(d) Now, add a dropout of 0.3 rate on the second fully connected layer. What is the impact of dropout on the performance? Provide the plot for training and test after each epoch. What happens if you decrease or increase the dropout rate?