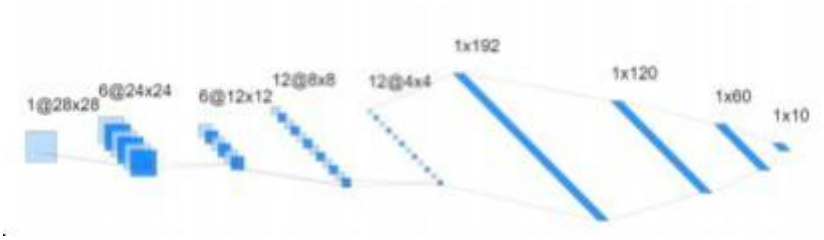
MCTE 7106 MACHINE VISION: QUIZ 3 (30 marks)
Name : Mohammad Najib Bin Abdul Rahim
Matric No: G1923437

One day while wandering around a clothing store at KL East Mall, you stumbled upon a pretty girl who is choosing a dress for Hari Raya. It turns out that the girl is visually impaired and had a hard time distinguishing between an abaya and a kebaya. To help people with the similar situation, you then decided to develop an AI system to identify the type of clothes using a Convolutional Neural Networks (ConvNet). In order to train the network, you decide to use the fashion MNIST dataset which is freely available on Pytorch.

1) [5 marks] Given the problem, what is the most appropriate loss function to use? Justify your answer.

2) [10 marks] Create and train a ConvNet corresponding to the following architecture:



For training, initialize your weights using the Xavier initialization, use ReLU as the activation functions, a learning rate of 0.1 with SGD optimizer. You will train your neural network for 50 epochs. What is the final train and test accuracy obtained? Provide accuracy and loss plots on the train and test set per each epoch. Looking at the loss through epochs, discuss what you observe.

3) [10 marks] 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 and 10. What do you observe? Explain.

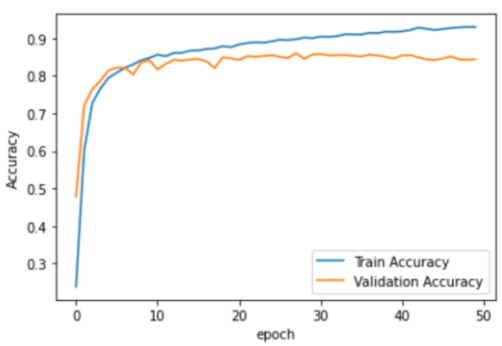
4) [5 marks] Now, add a dropout of 0.3 rate on the second fully connected layer. What is the effect of the dropout on the performance? Provide the plot for training and testing after each epoch. What happens if you decrease or increase the dropout rate?

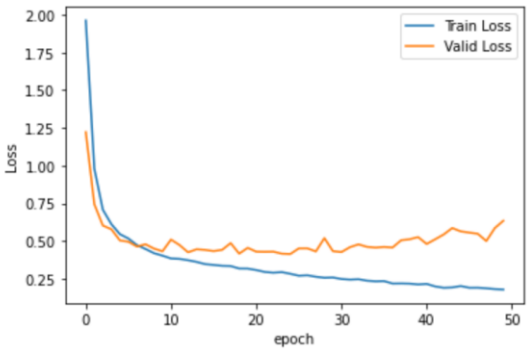
SOLUTION

1) Given the problem, what is the most appropriate loss function to use? Justify your answer.
   Cross entropy loss function has been selected for this classification problem. This type of loss function increases as the predicted probability diverges from the actual label. In other words, the most important justification is that cross entropy loss function penalizes heavily the predictions that are confident but wrong.

2) The final test (validation) and train loss and accuracy as follow:

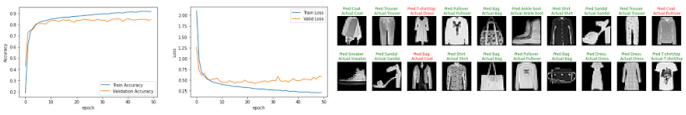
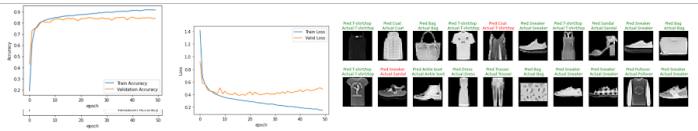
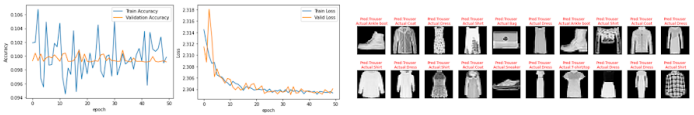
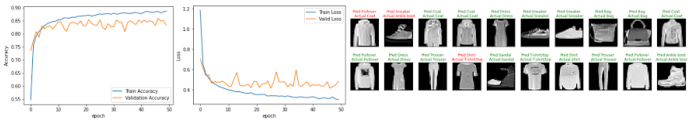
Epoch	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy
50	0.214532	0.588002	0.917417	0.844208





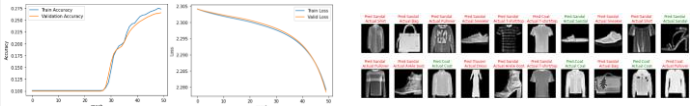
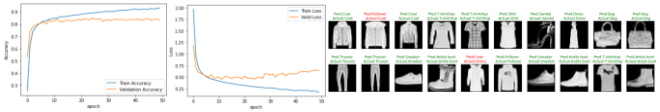
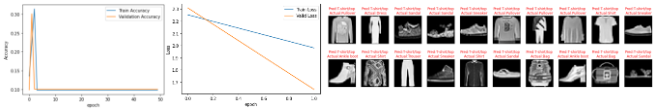
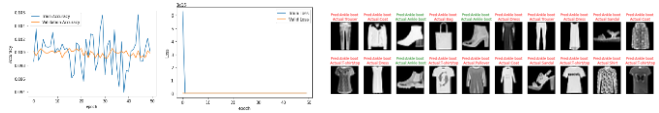
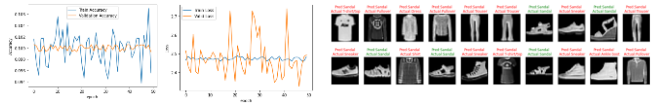
The train and validation loss diverges after about 9 epochs. This indicates further training did not improve the model. The training loss was lower than the validation loss indicates overfitting.

3) 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 and 10. What do you observe? Explain.

Question	Epoch	Weight Initialization	Optimizer	Activation Function	Learning Rate	Dropout	Results	Epoch	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy	Epoch Duration	Run Duration
2	50	Xavier	SGD	ReLU	0.1	-		50	0.214532	0.588002	0.917417	0.844208	15.165626	780.563304
3	50	Xavier	SGD	Tanh		-		50	0.152126	0.484116	0.943667	0.856563	15.587285	785.091051
3	50	Xavier	SGD	Sigmoid	0.1	-		50	2.303341	2.304094	0.099833	0.099146	16.026846	782.988121
3	50	Xavier	SGD	ELU		-		50	0.30117	0.480585	0.887333	0.833167	14.425601	727.247739

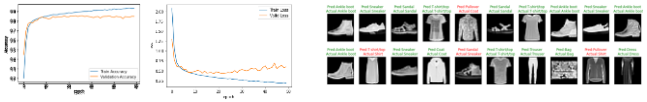
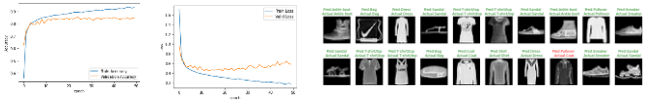
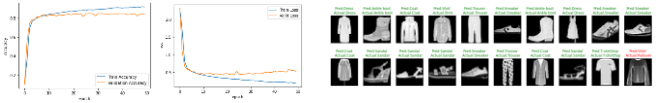
Based on the final train and validation loss and accuracy, the model performance can be ranked as follow from best to least performing in comparison of different activation functions:

1. Tanh
2. ReLU
3. ELU
4. Sigmoid

Question	Epoch	Weight Initialization	Optimizer	Activation Function	Learning Rate	Dropout	Results	Epoch	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy	Epoch Duration	Run Duration
3	50	Xavier	SGD	ReLU	0.001	-		50	2.279183	2.278521	0.272917	0.265104	16.027899	770.304392
3	50	Xavier	SGD		0.1	-		50	0.176385	0.640357	0.931417	0.832417	14.973531	755.360702
3	50	Xavier	SGD		0.5	-		50	NaN	NaN	0.09775	0.100562	14.089588	738.270756
3	50	Xavier	SGD		1	-		50	2.31E+00	2.307827	0.100167	0.099875	15.170194	725.653378
3	50	Xavier	SGD		10	-		50	2.484492	2.493118	0.095583	0.099312	20.287452	1049.019675

Keeping ReLU with 5 varying learning rate, it is observed that the optimum training and validation loss and accuracy recoded at learning rate 0.1. The rest of the learning rate resulted very poor training and validation loss and accuracy. Higher learning rate requires higher computing but poor losses and accuracy.

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

Question	Epoch	Weight Initialization	Optimizer	Activation Function	Learning Rate	Dropout	Results	Epoch	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy	Epoch Duration	Run Duration
4	50	Xavier	SGD	ReLU	0.1	0.3 @ FC#2		50	0.181733	0.591917	0.929	0.843583	21.089604	1063.014246
4	50	Xavier	SGD	ReLU	0.1	0.001 @ FC#2		50	0.177949	0.57844	0.932333	0.845021	14.583184	759.463539
4	50	Xavier	SGD	ReLU	0.1	1 @ FC#2		50	0.208661	0.528507	0.91925	0.844667	14.821173	752.838753

Dropout rate of 0.001 resulted the best training and validation loss and accuracy with acceptable run duration. Higher dropout rate of 0.3 took longer to compute but with no improvement in loss and accuracy. Higher dropout rate of 1.0 took shorter run duration with lower validation loss and similar validation accuracy as compared to 0.001. Hence, dropout rate of 1.0, learning rate 0.1, ReLU activation is the best performance among all in this exercise. However, the overall validation loss and accuracy cannot improve further without revisiting the CNN architecture.