CS6910 PG Section Instructor: Mitesh Khapra

Assignment 1

Name: Sandeep Kumar Suresh

Roll Number: EE23S059

WandB Report Link: wandb Report Link

GitHub Repository Link: Github Repository

CS6910 PG Section Instructor: Mitesh Khapra Assignment 1

Report for CS6910 Assignment 1

Sandeep Kumar Suresh ee23s059

- ***Github Link*** - Github

Question 1 (2 Marks)

Download the fashion-MNIST dataset and plot 1 sample image for each class as shown in the grid below. Use `from keras.datasets import fashion_mnist` for getting the fashion mnist dataset.

Question 2 (10 Marks)

Submitted the code in moodle

Question 3 (24 Marks)

Code Submitted in Moodle

Question 4 (10 Marks)

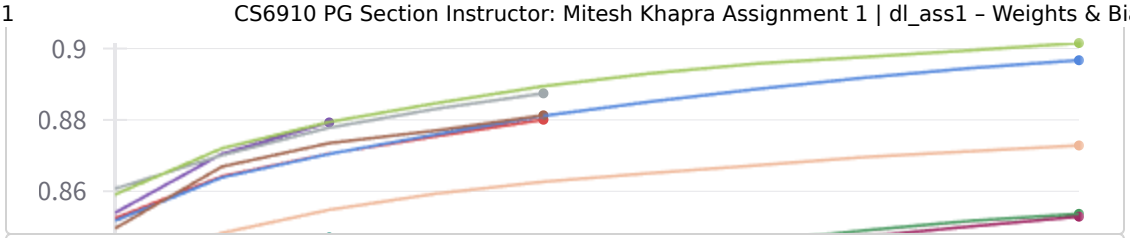
Wandb provides different 3 methods to do Hyperparameter Sweeping which are Random , bayes and grid .

Grid Search tries every combination of the hyperparameter

Random Search Select each new combination at random according to provided distributions

Bayesian Search creates a probabilistic model of metric score as a function of the hyperparameters, and choose parameters with high probability of improving the metric .

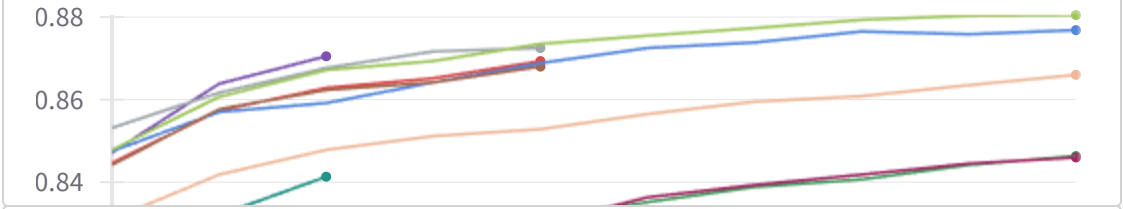
I have choosen bayesian Search as it is computatioally effective and uses a probabilistic approach to find the best set of hyperparameter that can maximize the Validation Accuracy



val_Accuracy

Showing first 10 runs

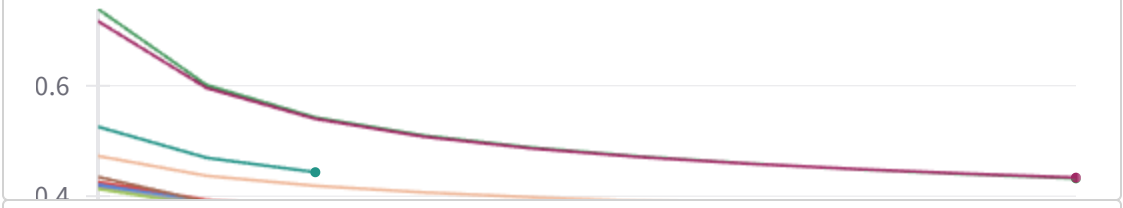
- hidden_layer:5_mini_batch_size:32_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:4_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:3_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:32_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1



val_loss

Showing first 10 runs

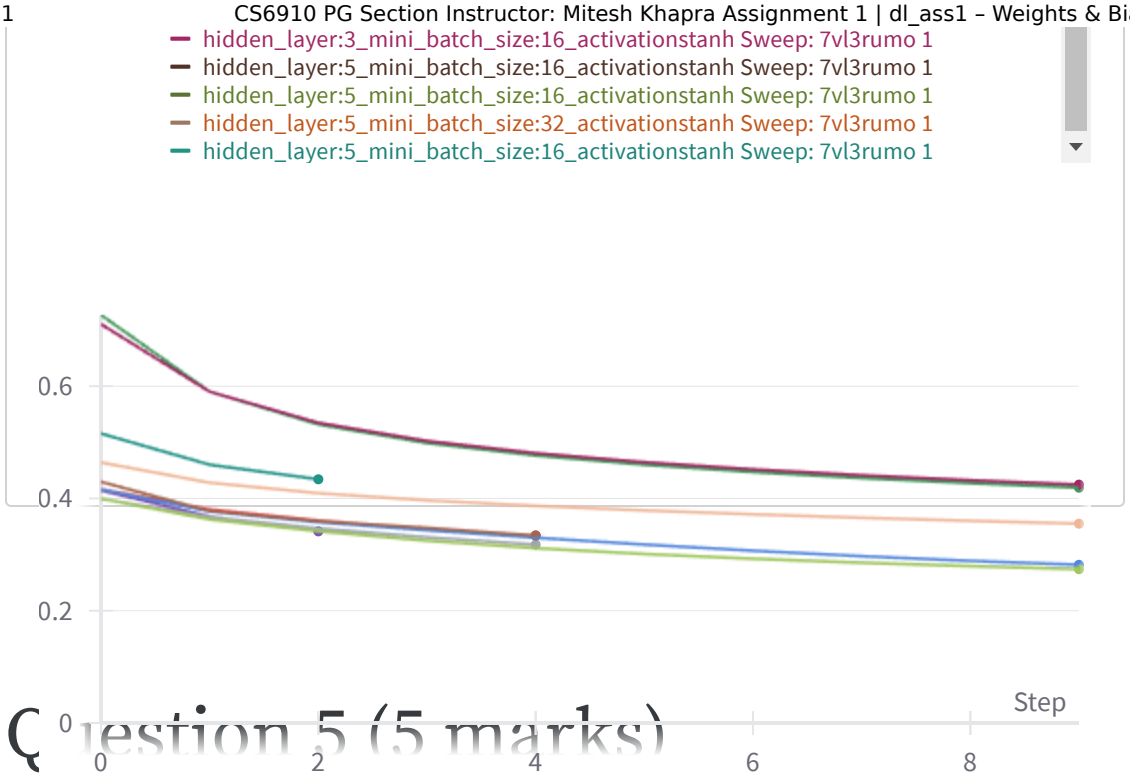
- hidden_layer:5_mini_batch_size:32_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:4_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:3_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:32_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1



train_Loss

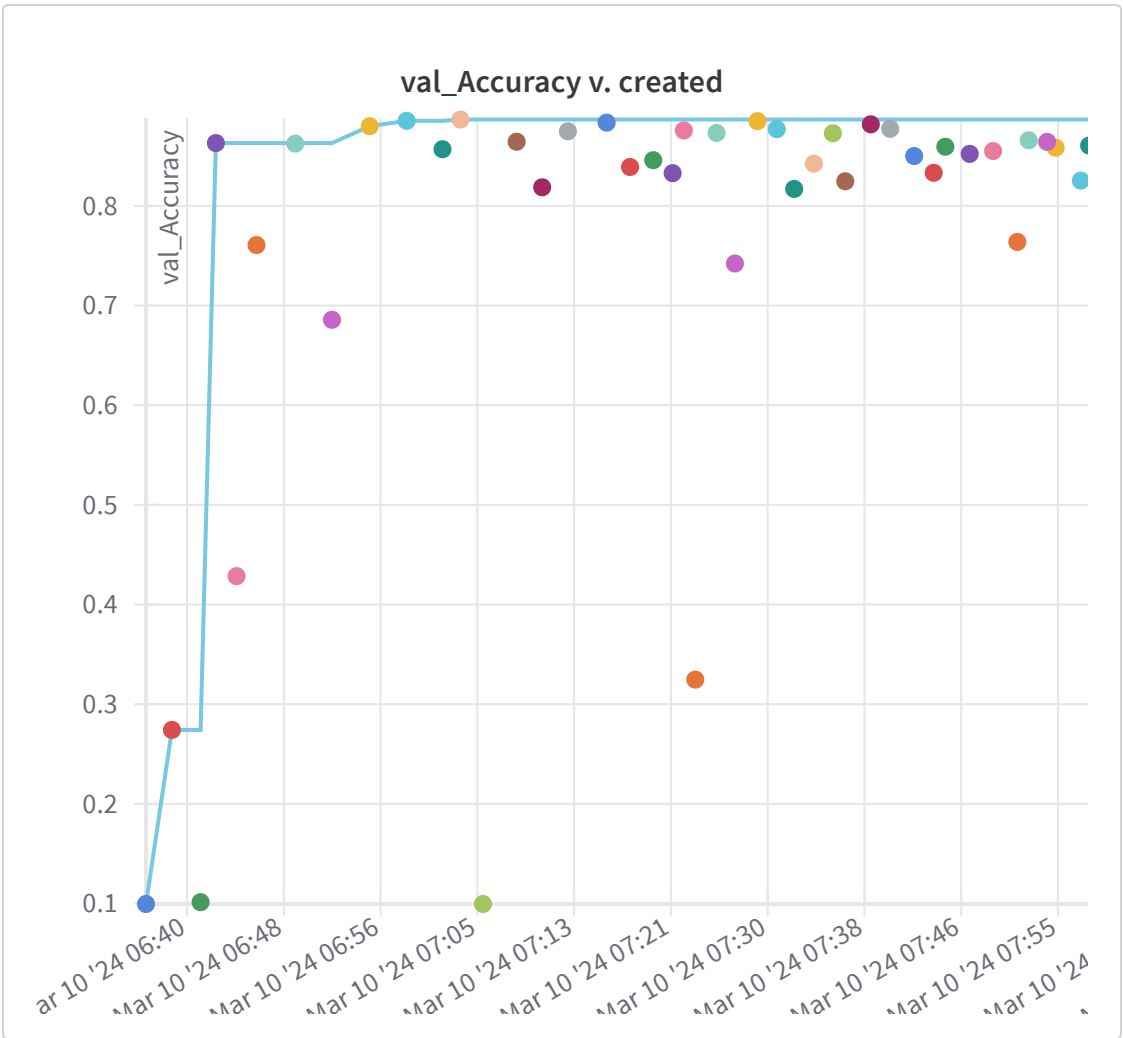
Showing first 10 runs

- hidden_layer:5_mini_batch_size:32_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:4_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1
- hidden_layer:5_mini_batch_size:16_activationstanh Sweep: 7vl3rumo 1



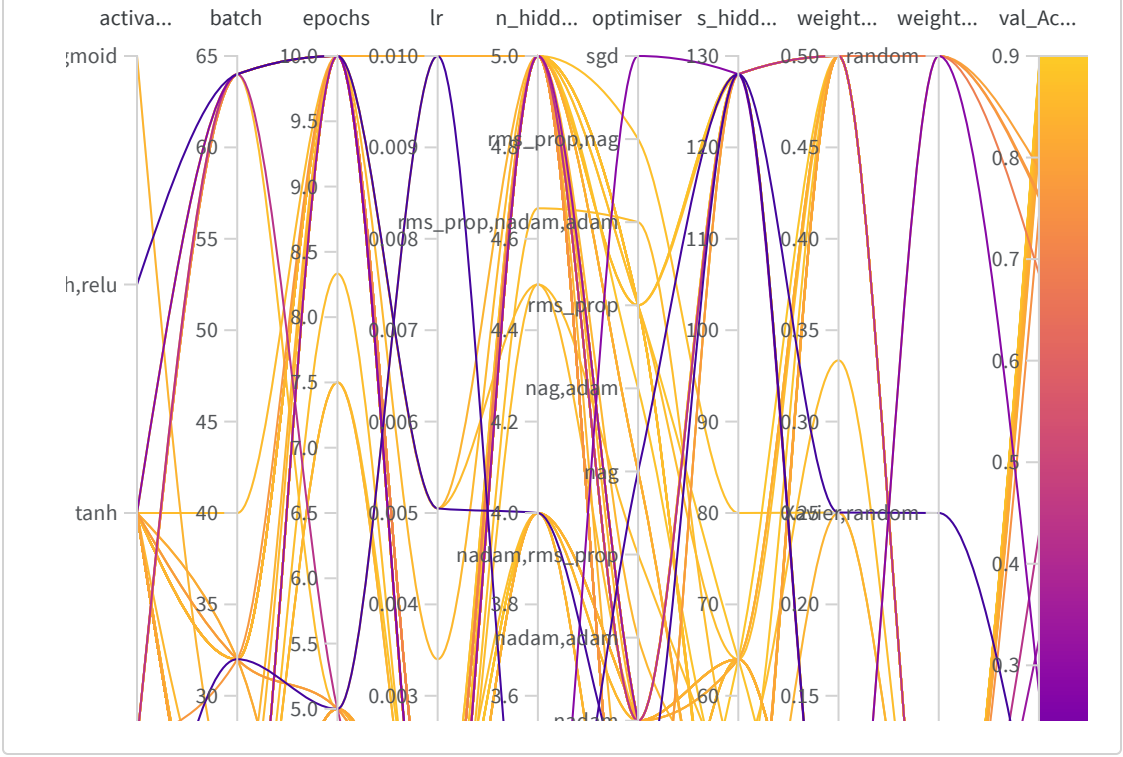
Question 5 (5 marks)

- I got a maximum Validation Accuracy of around **88.82 %.**
- The minimum Validation Accuracy is around **9.95 %**



Question 6 (20 Marks)

1. The highest validation Accuracy is **88.82% without Regularization** . The configuration was nadam,n_layers = 4 ,s_hidden_layer = 64 , batch_size = 32,lr = 1e-3,Xavier Initialization , epochs = 10 ,tanh activation.
2. The lowest Validation Accuracy Observed was around **9.95% without Regularization** . The configuration was mgd,n_layers = 5 ,s_hidden_layer = 128 , batch_size = 64,lr = 1e-2,Xavier Initialization , epochs = 10 ,tanh activation.
3. When I grouped the sweep based on Optimiser that gave the best validation Accuracy , I observed that most gave best result with Xavier Initialization.
4. When filtered with group of Validation Accuracy and Optimiser Together , nadam and adam came in the top - 20 list. Most of the activation function was tanh , learning rate of 1e-2 , hidden layer of 4 or 5 with Xavier Initialization was most common observation. This states the fact that nadam and adam are the best optimizer.(Note: nag and rms_prop came into the list).
5. Based on the Correlation Plot , n_hidden_layer is a parameter of importance which neccessarily had a positive correleation to maximize the validation Accuracy
6. nadam,sgd,nag,mgd were the optimizer which gave validation accuracy below 65% . Eventhough nadam gave highest validation , maybe the learning rate of 1e-4,sigmoid activation and size of hidden layer could have made it performance degrade.
7. Definitly nadam, adam can be used as optimizer configuration to get more than95% with Regularization and Data Augmentation Added. The below two plots are plots that does not involve Regularization.

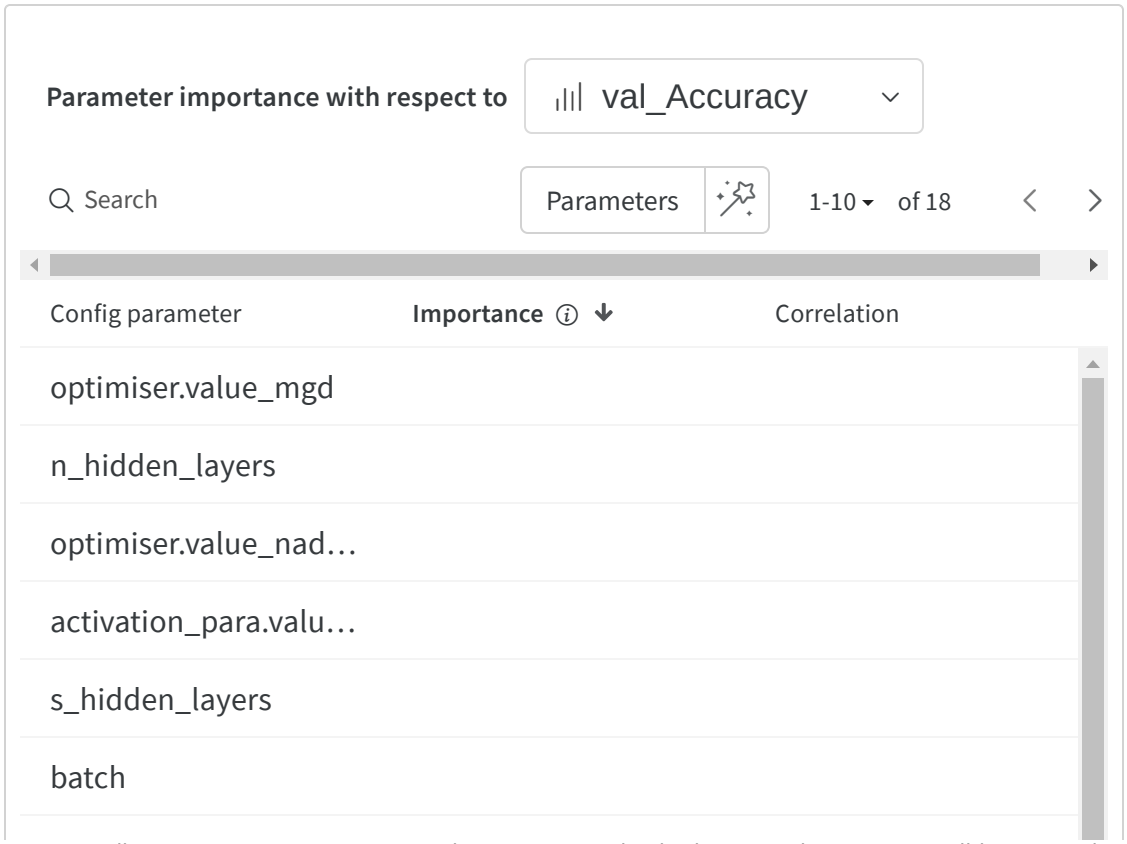


Val Accuracy with Optimizers

Showing first 10 bars

optimiser: nadam Sweep: 7vl3rumo 1

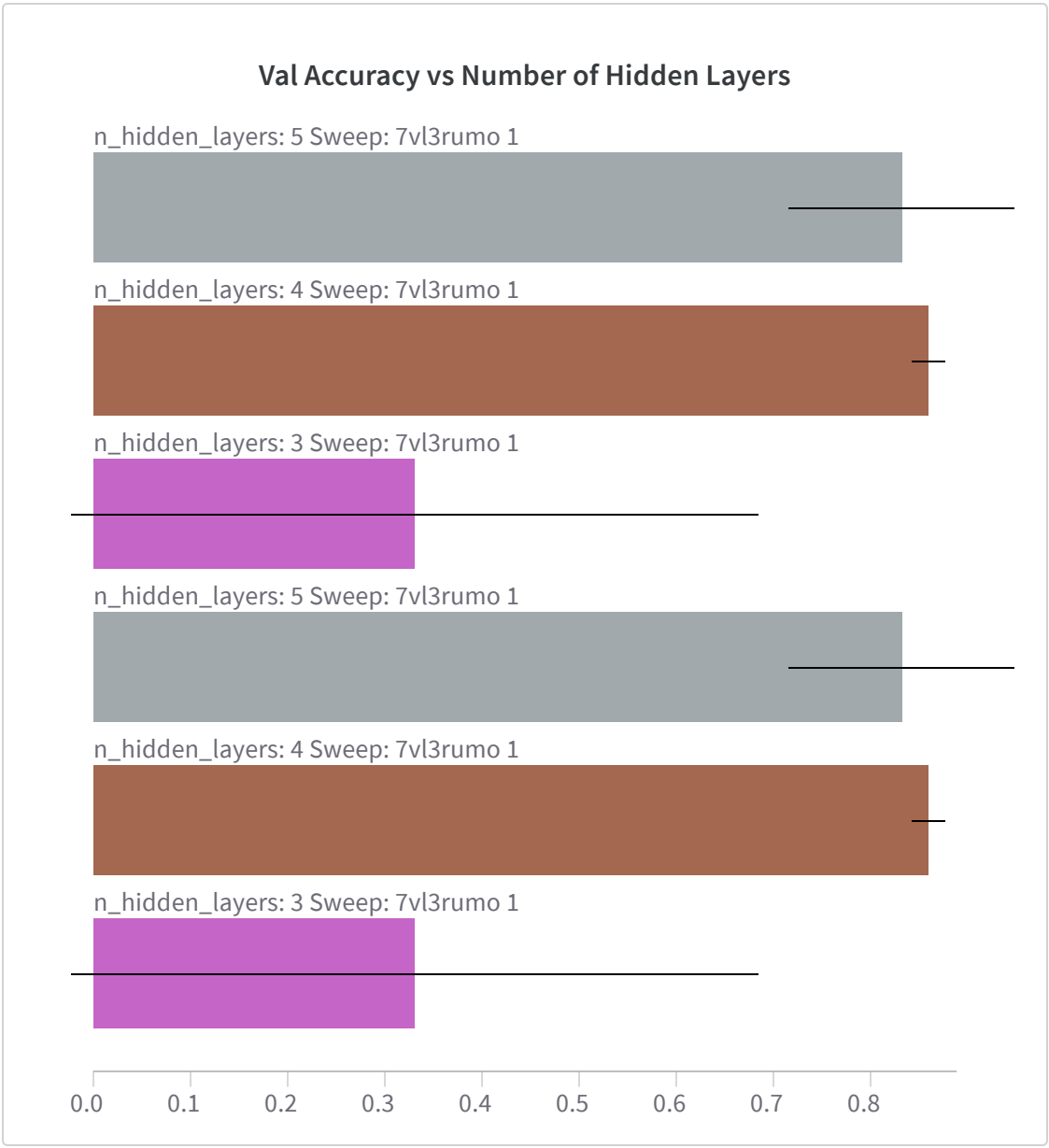
- Now based on the "Correlation Summary" , we can see that n_hidden_layers is having a positive correlation on the validation Accuracy . Therefore on increasing the n_hidden_layers in the network , the accuracy may increase.
- Tanh Activation based on the "Parallel co-ordinates plot" and a "correlation summary" have shown to increase the validation Accuracy . This has a higher positive correlation than n_hidden_layers
- The nadam optimizer as expected tends to have positive correlation . This could vary from different configuration used.
- Epoch has little importance and have a negative correlation. This could mean that increasing the number of epoch could be computationally expensive and may end in regions where the loss is very high.
- The adam optimizer tends to work for certain configuration based on "Correlation Summary" . But it is of least important. This is also the case with rms_prop.
- Xavier Initialization works best for all the parameter configuration



activation_para.valu…

optimiser.value_sgd

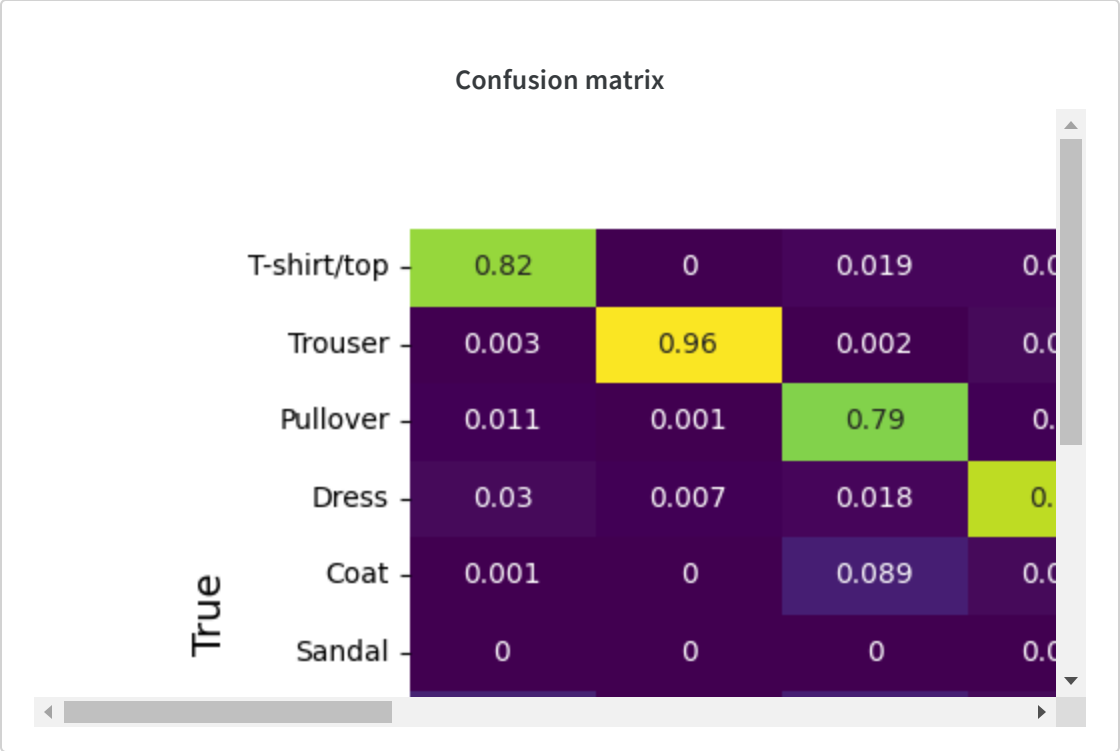
Runtime



Question 7 (10 Marks)

For the best model identified above, report the accuracy on the test set of fashion_mnist and plot the confusion matrix as shown below. More marks for creativity (less marks for producing the plot shown below as it is)

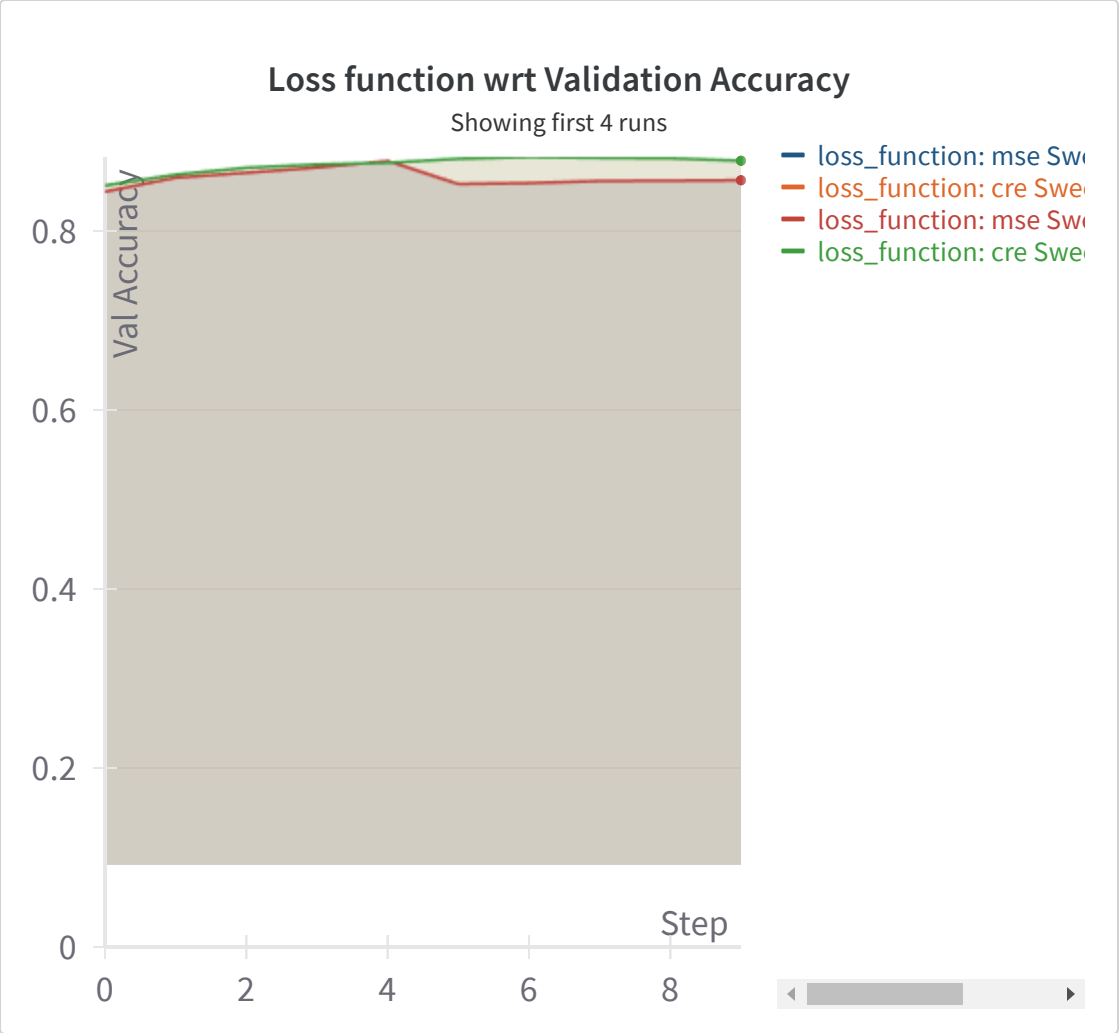
- The best model configuration identified is nadam with n_layers = 4 ,s_hidden_layer = 64 , batch_size = 32,lr = 1e-3,Xavier Initialization , epochs = 10 , cross-entropy loss and tanh activation.
- We can see that most shirt got mispredicted as T-shirt and some as Pullovers and Coats
- T-shirts , Pullovers and Coats are similar to each other . Therfore there are mispredictions for these classes.



Question 8 (5 Marks)

In all the models above you would have used cross entropy loss. Now compare the cross entropy loss with the squared error loss. I would again like to see some automatically generated plots or your own plots to convince me whether one is better than the other.

- I ran Sweep with cross entropy loss and squared error loss as parameters
- We can see that the cross entropy loss performed better than squared error loss wrt to Validation Accuracy.Hence we can say that cross entropy loss is better.



Question 9 (10 Marks)

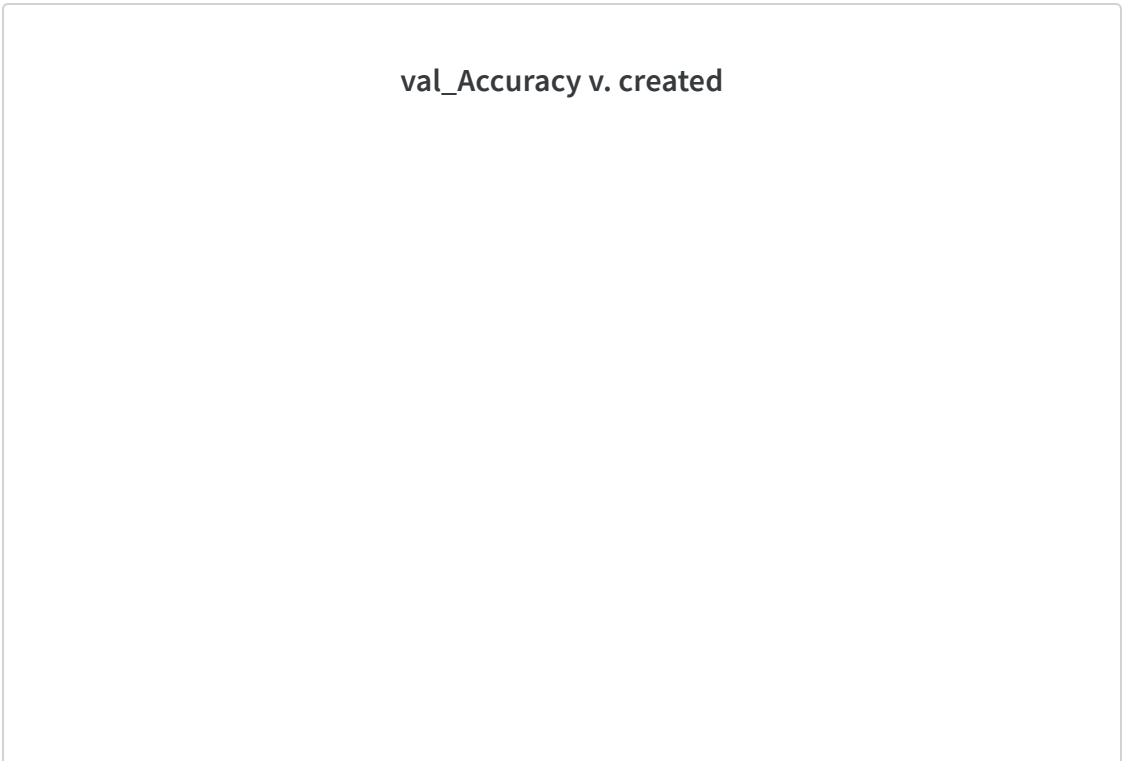
https://github.com/sandeepkumarsuresh/cs6910_assignment1

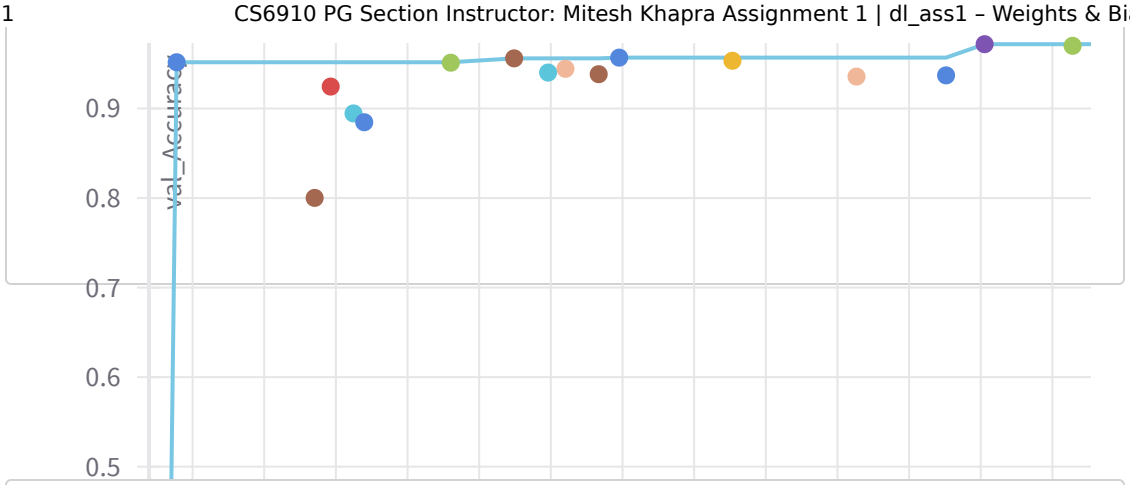
Question 10 (10 Marks)

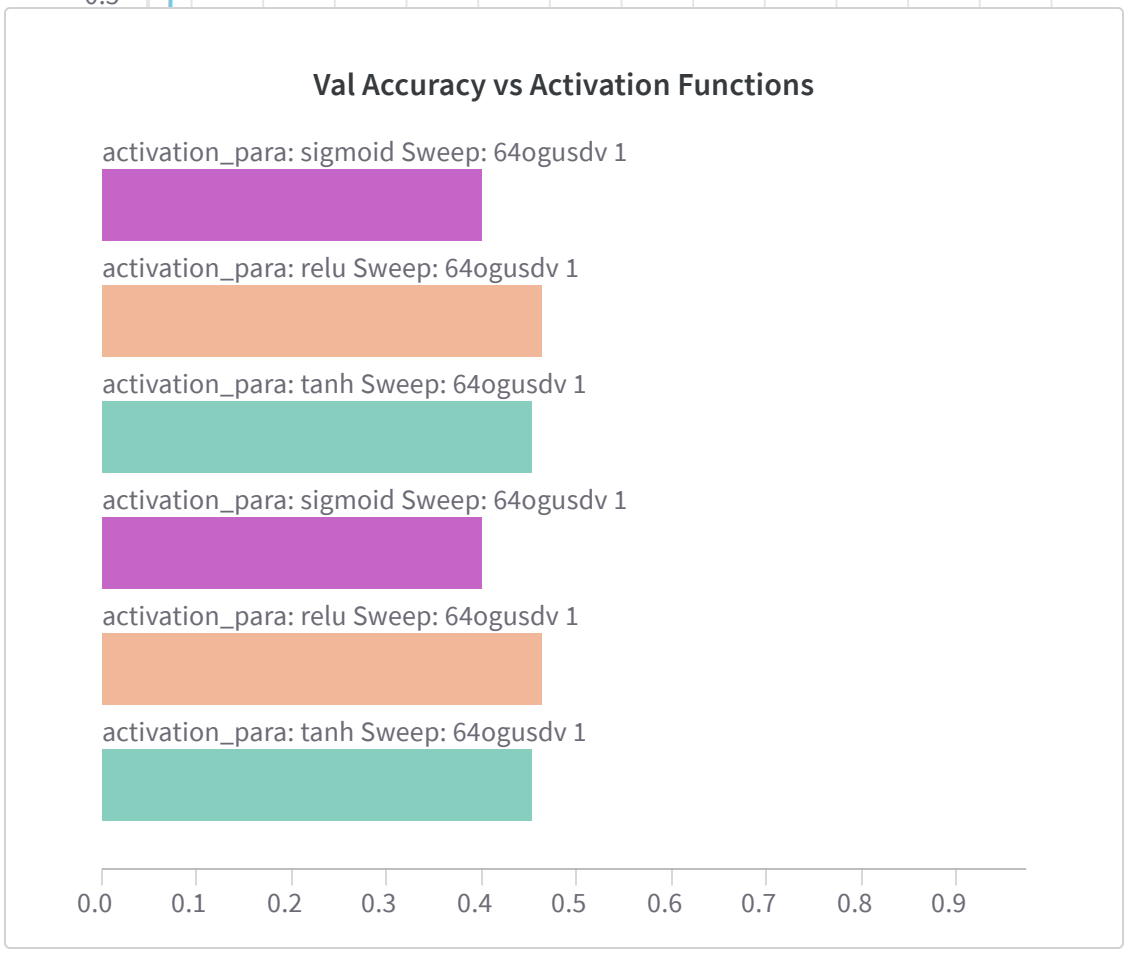
- Based on my understanding from the experimentation of Fashion MNIST Dataset . I would work on

1. Optimiser : Constraining it to adam, nadam and nag
2. Activation Function: tanh , relu or sigmoid
3. Number of Hidden Layer: 3,4,5

- The highest Validation Accuracy for MNIST Dataset is **97.31% .**
- The Configuration-1 for **97.31%** is adam optimiser,tanh activation,batch=16,epoch =5,lr=1e-3,Xavier Initialization,n_hidden layers =3 , s_hidden_layer =128
- The Configuration-2 for **97 %** nadam optimiser,tanh activation,batch=64,epoch =5,lr=1e-3,Xavier Initialization,n_hidden layers =4 , s_hidden_layer =128
- The Configuration-2 for **96.7%** nag optimiser,relu activation,batch=32,epoch =5,lr=1e-3,Xavier Initialization,n_hidden layers =4 , s_hidden_layer =128
- Interestingly we can see that the relu function also had a correlation in getting accuracy above 95%.
- From this experiment what I observed is that based on the data we have , certain activation function will provide better result.In the Fashion MNIST Dataset , tanh activation gave better result .







Self Declaration

I, Sandeep Kumar Suresh, swear on my honour that I have written the code and the report by myself and have not copied it from the internet or other students.

Created with ❤️ on Weights & Biases.

https://wandb.ai/ee23s059/dl_ass1/reports/CS6910-PG-Section-Instructor-Mitesh-Khapra-Assignment-1--Vmlldzo3MTc2NjEz